

Db 181 LVFFATSF 188

RESULT 62

F69260

nitrate ABC transporter, permease protein (nrtB-1) homolog - *Archaeoglobus fulgidus*

C;Species: *Archaeoglobus fulgidus*

C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 21-Jul-2000

C;Accession: F69260

R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus*.

A;Reference number: A69250; MUID:98049343; PMID:9389475

A;Accession: F69260

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-244 <KLE>

A;Cross-references: GB:AE001100; GB:AE000782; NID:g2689423; PIDN:AAB91143.1;

PID:g2650562; TIGR:AF0086

C;Superfamily: *Synechococcus* nitrate transport protein nrtB

Query Match	70.7%;	Score 29;	DB 2;	Length 244;
Best Local Similarity	62.5%;	Pred. No. 1.4e+02;		
Matches	5;	Conservative	2;	Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8

::|||||

Db 186 VLFFAESF 193

RESULT 63

B86301

hypothetical protein F19K19.10 - *Arabidopsis thaliana*

C;Species: *Arabidopsis thaliana* (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001

C;Accession: B86301

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.; Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.; Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.; Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu,

S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda, M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken, S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A;Reference number: A86141; MUID:21016719; PMID:11130712  
 A;Accession: B86301  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-247 <STO>  
 A;Cross-references: GB:AE005172; NID:g9989059; PIDN:AAG10822.1; GSPDB:GN00141  
 C;Genetics:  
 A;Map position: 1

Query Match 70.7%; Score 29; DB 2; Length 247;  
 Best Local Similarity 71.4%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 LVFFAED 7  
 :|||:|  
 Db 169 VVFFSED 175

# RESULT 64

C75200  
 hypothetical protein PAB2263 - Pyrococcus abyssi (strain Orsay)  
 C;Species: Pyrococcus abyssi  
 C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
 C;Accession: C75200  
 R;anonymous, Genoscope  
 submitted to the EMBL Data Library, July 1999  
 A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.  
 A;Reference number: A75001  
 A;Accession: C75200  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-251 <KAW>  
 A;Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49050.1; PID:g5457559  
 A;Experimental source: strain Orsay  
 C;Genetics:  
 A;Gene: PAB2263  
 C;Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0225

Query Match 70.7%; Score 29; DB 2; Length 251;  
 Best Local Similarity 62.5%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 LVFFAEDF 8  
 :|| |||:  
 Db 111 VVFLAEDY 118

RESULT 65

T06661

hypothetical protein T6G15.140 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 22-Oct-1999

C;Accession: T06661

R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schueller, C.

submitted to the Protein Sequence Database, April 1999

A;Reference number: Z15791

A;Accession: T06661

A;Molecule type: DNA

A;Residues: 1-273 <BEV>

A;Cross-references: EMBL:AL049656; GSPDB:GN00062; ATSP:T6G15.140

A;Experimental source: cultivar Columbia; BAC clone T6G15

C;Genetics:

A;Gene: ATSP:T6G15.140

A;Map position: 4

A;Introns: 48/1; 95/3; 111/3; 144/1; 220/3; 237/3

Query Match 70.7%; Score 29; DB 2; Length 273;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAE 6  
|||||  
Db 216 LVFFAE 221

RESULT 66

F86805

cation transporter yogJ [imported] - Lactococcus lactis subsp. lactis (strain IL1403)

C;Species: Lactococcus lactis subsp. lactis

C;Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001

C;Accession: F86805

R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich, S.D.; Sorokin, A.

Genome Res. 11, 731-753, 2001

A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis IL1403.

A;Reference number: A86625; MUID:21235186; PMID:11337471

A;Accession: F86805

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-314 <STO>

A;Cross-references: GB:AE005176; PID:g12724436; PIDN:AAK05544.1; GSPDB:GN00146

A;Experimental source: strain IL1403

C;Genetics:

A;Gene: yogJ

Query Match 70.7%; Score 29; DB 2; Length 314;  
Best Local Similarity 85.7%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
|| ||||

Db 163 LVIFAED 169

RESULT 67

H71729

hypothetical protein RP189 - Rickettsia prowazekii

C;Species: Rickettsia prowazekii

C;Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000

C;Accession: H71729

R;Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark, U.C.M.; Podowski, R.M.; Naeslund, A.K.; Eriksson, A.S.; Winkler, H.H.; Kurland, C.G.

Nature 396, 133-140, 1998

A;Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.

A;Reference number: A71630; MUID:99039499; PMID:9823893

A;Accession: H71729

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-321 <AND>

A;Cross-references: GB:AJ235270; GB:AJ235269; NID:g3860572; PIDN:CAA14655.1;

PID:e1342498; PID:g3860754; GSPDB:GN00081

A;Experimental source: strain Madrid E

C;Genetics:

A;Gene: RP189

C;Superfamily: Rickettsia prowazekii hypothetical protein RP189

Query Match	70.7%;	Score 29;	DB 2;	Length 321;
Best Local Similarity	71.4%;	Pred. No. 1.8e+02;		
Matches	5;	Conservative	1;	Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAED 7

|:||| |

Db 179 LIFFAHD 185

RESULT 68

T25676

hypothetical protein F08D12.8 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C;Accession: T25676

R;Le, T.; Waterston, R.

submitted to the EMBL Data Library, December 1996

A;Description: The sequence of C. elegans cosmid F08D12.

A;Reference number: Z20068

A;Accession: T25676

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-332 <LET>

A;Cross-references: EMBL:U80840; PIDN:AAB37932.1; GSPDB:GN00020; CESP:F08D12.8

A;Experimental source: strain Bristol N2; clone F08D12

C;Genetics:

A;Gene: CESP:F08D12.8

A;Map position: 2

A;Introns: 27/2



Query Match 70.7%; Score 29; DB 2; Length 332;  
Best Local Similarity 62.5%; Pred. No. 1.9e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
||||:|:  
Db 73 LVFFKKDY 80

RESULT 69

S35229

hypD' protein - Bradyrhizobium japonicum (fragment)

C;Species: Bradyrhizobium japonicum

C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 14-Apr-2003

C;Accession: S35229; S28641

R;van Soom, C.; Verreth, C.; Sampaio, M.J.; Vanderleyden, J.

Mol. Gen. Genet. 239, 235-240, 1993

A;Title: Identification of a potential transcriptional regulator of hydrogenase activity in free-living Bradyrhizobium japonicum strains.

A;Reference number: S35229; MUID:93287991; PMID:8510650

A;Accession: S35229

A;Molecule type: DNA

A;Residues: 1-347 <VAN>

A;Cross-references: EMBL:Z17373; NID:g39534; PIDN:CAA78988.1; PID:g39535

C;Genetics:

A;Gene: hypD'

C;Superfamily: hydrogenase maturation factor, HypD type

Query Match 70.7%; Score 29; DB 2; Length 347;  
Best Local Similarity 83.3%; Pred. No. 2e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 FFAEDF 8  
||||:|  
Db 182 FFAEEF 187

RESULT 70

S59678

HST2 protein - yeast (Saccharomyces cerevisiae)

N;Alternate names: protein LPA2c; protein YPL015c

C;Species: Saccharomyces cerevisiae

C;Date: 13-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 29-Oct-1999

C;Accession: S59678

R;Hall, J.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.K.; Vo, D.H.; Wang, Y.; Winnett, E.

submitted to the EMBL Data Library, August 1995

A;Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.

A;Reference number: S59677

A;Accession: S59678

A;Molecule type: DNA

A;Residues: 1-357 <HAL>

A;Cross-references: EMBL:U33335; NID:g965076; PIDN:AAB68090.1; PID:g965078;

MIPS:YPL015c

C;Genetics:

A;Gene: SGD:HST2

A;Cross-references: SGD:S0005936; MIPS:YPL015c

A;Map position: 16L  
C;Keywords: transmembrane protein  
F;218-234/Domain: transmembrane #status predicted <TMM>

Query Match 70.7%; Score 29; DB 2; Length 357;  
Best Local Similarity 71.4%; Pred. No. 2e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
:||| ||  
Db 181 IVFFGED 187

RESULT 71  
S32877

hypD protein - Rhizobium leguminosarum bv. viciae  
C;Species: Rhizobium leguminosarum bv. viciae  
C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 14-Apr-2003  
C;Accession: S32877  
R;Rey, L.; Murillo, J.; Hernando, Y.; Hidalgo, E.; Cabrera, E.; Imperial, J.; Ruiz-Argueeso, T.  
Mol. Microbiol. 8, 471-481, 1993  
A;Title: Molecular analysis of a microaerobically induced operon required for hydrogenase synthesis in Rhizobium leguminosarum biovar viciae.  
A;Reference number: S32873; MUID:93316844; PMID:8326860  
A;Accession: S32877  
A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 1-385 <REY>  
A;Cross-references: EMBL:X52974; NID:g1167855; PIDN:CAA37163.1; PID:g48736  
C;Genetics:  
A;Gene: hypD  
C;Superfamily: hydrogenase maturation factor, HypD type

Query Match 70.7%; Score 29; DB 2; Length 385;  
Best Local Similarity 83.3%; Pred. No. 2.2e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 FFAEDF 8  
||||:|  
Db 215 FFAEEF 220

RESULT 72  
T25049

hypothetical protein T21B4.1 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: T25049  
R;Smye, R.  
submitted to the EMBL Data Library, October 1996  
A;Reference number: Z19974  
A;Accession: T25049  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-394 <WIL>  
A;Cross-references: EMBL:Z81124; PIDN:CAB03370.1; GSPDB:GN00020; CESP:T21B4.1

A;Experimental source: clone T21B4  
C;Genetics:  
A;Gene: CESP:T21B4.1  
A;Map position: 2  
A;Introns: 33/1; 57/1; 105/1; 183/3; 312/3

Query Match 70.7%; Score 29; DB 2; Length 394;  
Best Local Similarity 62.5%; Pred. No. 2.3e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
|:| | |  
Db 226 LIFFLPDF 233

RESULT 73

E90553

hypothetical protein MYPU\_3330 [imported] - Mycoplasma pulmonis (strain UAB CTIP)

C;Species: Mycoplasma pulmonis

C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001

C;Accession: E90553

R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.;

Moszer, I.; Dybvig, K.; Wroblewski, H.; Viari, A.; Rocha, E.P.C.; Blanchard, A.

Nucleic Acids Res. 29, 2145-2153, 2001

A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis.

A;Reference number: A99512; MUID:21267165; PMID:11353084

A;Accession: E90553

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-420 <KUR>

A;Cross-references: GB:AL445566; PID:g14089747; PIDN:CAC13506.1; GSPDB:GN00153

A;Experimental source: strain UAB CTIP

C;Genetics:

A;Gene: MYPU\_3330

A;Genetic code: SGC3

C;Superfamily: histidine-tRNA ligase; histidine-tRNA ligase homology

Query Match 70.7%; Score 29; DB 2; Length 420;  
Best Local Similarity 50.0%; Pred. No. 2.4e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
::|||::|  
Db 141 IIFFAKNF 148

RESULT 74

E71082

probable phospho-sugar mutase 2 - Pyrococcus horikoshii

C;Species: Pyrococcus horikoshii

C;Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 20-Jun-2000

C;Accession: E71082

R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto,

S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura,

K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka,

T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H. DNA Res. 5, 55-76, 1998

A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3.

A;Reference number: A71000; MUID:98344137; PMID:9679194

A;Accession: E71082

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-455 <KAW>

A;Cross-references: GB:AP000004; NID:g3236131; PIDN:BAA30019.1; PID:g3257336

A;Experimental source: strain OT3

A;Note: this accession replaces an interim accession for a sequence replaced by GenBank

C;Genetics:

A;Gene: PH0923

C;Superfamily: phosphomannomutase

Query Match 70.7%; Score 29; DB 2; Length 455;  
Best Local Similarity 71.4%; Pred. No. 2.6e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
: || |||  
Db 131 LFFKEDF 137

#### RESULT 75

B84230

hypothetical protein Vng0727c [imported] - *Halobacterium* sp. NRC-1

C;Species: *Halobacterium* sp. NRC-1

C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C;Accession: B84230

R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.R.; Baliga, N.; Thorsson, V.; Sbrogna, J.; Swartzell, S.; Weir, D.; Hall, J.; Dahl, T.A.; Welte, R.; Goo, Y.A.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablonski, P.E.; Krebs, M.P.; Angevine, C.M.; Dale, H.; Isenbarger, T.A.; Peck, R.F.; Pohlschrod, M.; Spudich, J.L.; Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Liang, P.; Riley, M.; Hood, L.; DasSarma, S.

A;Title: Genome sequence of *Halobacterium* species NRC-1.

A;Reference number: A84160; MUID:20504483; PMID:11016950

A;Accession: B84230

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-494 <STO>

A;Cross-references: GB:AE004437; NID:g10580306; PIDN:AAG19206.1; GSPDB:GN00138

C;Genetics:

A;Gene: VNG0727C

Query Match 70.7%; Score 29; DB 2; Length 494;  
Best Local Similarity 83.3%; Pred. No. 2.9e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 FFAEDF 8

Db            || |||  
             125 FFVEDE 130

RESULT 76

T36945

hypothetical protein SCJ1.12 - Streptomyces coelicolor

C;Species: Streptomyces coelicolor

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C;Accession: T36945

R;Seeger, K.J.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.;  
Rajandream, M.A.

submitted to the EMBL Data Library, September 1999

A;Reference number: Z21607

A;Accession: T36945

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-508 <SEE>

A;Cross-references: EMBL:AL109962; PIDN:CAB53130.1; GSPDB:GN00070;

SCOEDB:SCJ1.12

A;Experimental source: strain A3(2)

C;Genetics:

A;Gene: SCOEDB:SCJ1.12

Query Match                    70.7%;   Score 29;   DB 2;   Length 508;  
Best Local Similarity       71.4%;   Pred. No. 2.9e+02;  
Matches       5;   Conservative       1;   Mismatches       1;   Indels       0;   Gaps       0;

Qy            1 LVFFAED 7  
             :||| ||  
Db            25 MVFFVED 31

RESULT 77

T11391

cytochrome-c oxidase (EC 1.9.3.1) chain I - fat dormouse mitochondrion

C;Species: mitochondrion Myoxus glis (fat dormouse)

C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 24-Oct-2000

C;Accession: T11391

R;Reyes, A.; Pesole, G.; Saccone, C.

Mol. Biol. Evol. 15, 499-505, 1998

A;Title: Complete mitochondrial DNA sequence of the fat dormouse, Glis glis:  
further evidence of rodent paraphyly.

A;Reference number: Z17267; MUID:98242079; PMID:9580978

A;Accession: T11391

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-513 <REY>

A;Cross-references: EMBL:AJ001562; NID:g3127895; PIDN:CAA04832.1; PID:g3127898;  
GSPDB:GN00132

C;Genetics:

A;Gene: COI

A;Genome: mitochondrion

A;Genetic code: SGCl

C;Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I  
homology

C;Keywords: copper; electron transfer; magnesium; membrane-associated complex; metalloprotein; mitochondrion; oxidoreductase; respiratory chain  
F;11-457/Domain: cytochrome-c oxidase chain I homology <CO1>  
F;240,290,291/Binding site: copper (His) #status predicted  
F;240-244/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted  
F;244/Binding site: oxygen (Tyr) #status predicted  
F;368/Binding site: magnesium (His) (shared with chain II) #status predicted

Query Match 70.7%; Score 29; DB 2; Length 513;  
Best Local Similarity 62.5%; Pred. No. 3e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
| || :||  
Db 423 LTFFPQDF 430

RESULT 78

G86587

heat shock protein-60 [imported] - Chlamydophila pneumoniae (strain J138)

C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001

C;Accession: G86587

R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.;

Shiba, T.; Ishii, K.; Hattori, M.; Kuhara, S.; Nakazawa, T.

Nucleic Acids Res. 28, 2311-2314, 2000

A;Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.

A;Reference number: A86491; MUID:20330349; PMID:10871362

A;Accession: G86587

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-526 <STO>

A;Cross-references: GB:BA000008; NID:g8979150; PIDN:BAA98985.1; GSPDB:GN00142

A;Experimental source: strain J138

C;Genetics:

A;Gene: groEL\_2

C;Superfamily: chaperonin groEL

Query Match 70.7%; Score 29; DB 2; Length 526;  
Best Local Similarity 75.0%; Pred. No. 3e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
|| ||||  
Db 245 LVIIAEDF 252

RESULT 79

D72036

heat shock protein-60 - Chlamydophila pneumoniae (strain CWL029)

C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae

C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000

C;Accession: D72036

R;Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.;

Grimwood, J.; Davis, R.W.; Stephens, R.S.

Nature Genet. 21, 385-389, 1999

A;Title: Comparative genomes of Clamydia pneumoniae and C. trachomatis.

A;Reference number: A72000; MUID:99206606; PMID:10192388  
A;Accession: D72036  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-526 <ARN>  
A;Cross-references: GB:AE001659; GB:AE001363; NID:g4377081; PIDN:AAD18915.1;  
PID:g4377086  
A;Experimental source: strain CWL029  
C;Genetics:  
A;Gene: groEL\_2  
C;Superfamily: chaperonin groEL

Query Match 70.7%; Score 29; DB 2; Length 526;  
Best Local Similarity 75.0%; Pred. No. 3e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
|| ||||  
Db 245 LVIIAEDF 252

RESULT 80

F81504

60 kDa chaperonin CP1095 [imported] - Chlamydophila pneumoniae (strain AR39)

C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae

C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-May-2000

C;Accession: F81504

R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.;  
Hickey, E.K.; Peterson, J.; Utterback, T.; Berry, K.; Bass, S.; Linher, K.;  
Weidman, J.; Khouri, H.; Craven, B.; Bowman, C.; Dodson, R.; Gwinn, M.; Nelson,  
W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, S.L.; Eisen, J.; Fraser,  
C.M.

Nucleic Acids Res. 28, 1397-1406, 2000

A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae  
AR39.

A;Reference number: A81500; MUID:20150255; PMID:10684935

A;Accession: F81504

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-526 <REA>

A;Cross-references: GB:AE002265; GB:AE002161; NID:g7189999; PIDN:AAF38865.1;

PID:g7190008; GSPDB:GN00122; TIGR:CP1095

A;Experimental source: strain AR39, HL cells

C;Genetics:

A;Gene: CP1095

C;Superfamily: chaperonin groEL

Query Match 70.7%; Score 29; DB 2; Length 526;  
Best Local Similarity 75.0%; Pred. No. 3e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
|| ||||  
Db 245 LVIIAEDF 252

RESULT 81

E82269  
 conserved hypothetical protein VC0880 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1)  
 C;Species: *Vibrio cholerae*  
 C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
 C;Accession: E82269  
 R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Umayam, L.A.; Gill, S.R.; Nelson, K.E.; Read, T.D.; Tettelin, H.; Richardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.; McDonald, L.; Utterback, T.; Fleishmann, R.D.; Nierman, W.C.; White, O.; Salzberg, S.L.; Smith, H.O.; Colwell, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 406, 477-483, 2000  
 A;Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
 A;Reference number: A82035; MUID:20406833; PMID:10952301  
 A;Accession: E82269  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-534 <HEI>  
 A;Cross-references: GB:AE004172; GB:AE003852; NID:g9655323; PIDN:AAF94042.1; GSPDB:GN00126; TIGR:VC0880  
 A;Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C;Genetics:  
 A;Gene: VC0880  
 A;Map position: 1

Query Match 70.7%; Score 29; DB 2; Length 534;  
 Best Local Similarity 75.0%; Pred. No. 3.1e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 ||||| |  
 Db 373 LVFFAAQF 380

RESULT 82  
 T32523  
 hypothetical protein C49A9.2 - *Caenorhabditis elegans*  
 C;Species: *Caenorhabditis elegans*  
 C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C;Accession: T32523  
 R;Fulton, B.; Wohldmann, P.  
 submitted to the EMBL Data Library, December 1997  
 A;Description: The sequence of *C. elegans* cosmid C49A9.  
 A;Reference number: Z21184  
 A;Accession: T32523  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-572 <FUL>  
 A;Cross-references: EMBL:AF036693; PIDN:AAB88335.1; GSPDB:GN00022; CESP:C49A9.2  
 A;Experimental source: strain Bristol N2; clone C49A9  
 C;Genetics:  
 A;Gene: CESP:C49A9.2  
 A;Map position: 4  
 A;Introns: 28/2; 83/3; 127/2; 186/3; 231/3; 281/3; 321/3; 363/2; 449/2; 530/3



Query Match 70.7%; Score 29; DB 2; Length 572;  
Best Local Similarity 75.0%; Pred. No. 3.3e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
||| |||  
Db 99 LVFSTEDF 106

RESULT 83

T24172

hypothetical protein R11A8.4 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C;Accession: T24172

R;Cummings, P.

submitted to the EMBL Data Library, March 1996

A;Reference number: Z19849

A;Accession: T24172

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-607 <WIL>

A;Cross-references: EMBL:Z70310; PIDN:CAA94364.1; GSPDB:GN00022; CESP:R11A8.4

A;Experimental source: clone R11A8

C;Genetics:

A;Gene: CESP:R11A8.4

A;Map position: 4

A;Introns: 18/3; 68/3; 111/3; 206/3; 500/3; 565/3

Query Match 70.7%; Score 29; DB 2; Length 607;  
Best Local Similarity 71.4%; Pred. No. 3.5e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
: ||| ||  
Db 298 IVFFGED 304

RESULT 84

S56284

hypothetical protein YFR029w - yeast (*Saccharomyces cerevisiae*)

N;Alternate names: hypothetical protein F001

C;Species: *Saccharomyces cerevisiae*

C;Date: 02-Sep-1995 #sequence\_revision 19-Oct-1995 #text\_change 20-Jun-2000

C;Accession: S56284; S62240; S63834

R;Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sasanuma, M.; Tsuchiya, Y.; Soeda, E.; Yokoyama, K.; Yamazaki, M.; Tashiro, H.; Eki, T.

submitted to the EMBL Data Library, May 1995

A;Description: Analysis of the nucleotide sequence of chromosome VI from *Saccharomyces cerevisiae*.

A;Reference number: S56186

A;Accession: S56284

A;Molecule type: DNA

A;Residues: 1-678 <MUR>

A;Cross-references: EMBL:D50617; NID:g836685; PIDN:BAA09268.1; PID:g836784; MIPS:YFR029w

R;Murakami, Y.  
submitted to the EMBL Data Library, December 1994  
A;Reference number: S62230  
A;Accession: S62240  
A;Molecule type: DNA  
A;Residues: 1-678 <MUW>  
A;Cross-references: EMBL:D44602; NID:g893419; PIDN:BAA08040.1; PID:g893424  
R;Eki, T.; Naitou, M.; Hagiwara, H.; Abe, M.; Ozawa, M.; Sasanuma, S.I.;  
Sasanuma, M.; Tsuchiya, Y.; Shibata, T.; Watanabe, K.; Ono, A.; Yamazaki, M.A.;  
Tashiro, H.; Hanaoka, F.; Murakami, Y.  
Yeast 12, 177-190, 1996  
A;Title: Fifteen open reading frames in a 30.8 kb region of the right arm of  
chromosome VI from *Saccharomyces cerevisiae*.  
A;Reference number: S63830; MUID:96287654; PMID:8686381  
A;Accession: S63834  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-678 <EKI>  
A;Cross-references: EMBL:D44602; NID:g893419; PIDN:BAA08040.1; PID:g893424  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, July  
1995  
C;Genetics:  
A;Gene: SGD:PTR3  
A;Cross-references: SGD:S0001925; MIPS:YFR029w  
A;Map position: 6R  
C;Superfamily: *Saccharomyces cerevisiae* hypothetical protein YFR029w

Query Match 70.7%; Score 29; DB 2; Length 678;  
Best Local Similarity 83.3%; Pred. No. 4e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 FFAEDF 8  
|| |||  
Db 654 FFVEDE 659

# RESULT 85

AF2275  
cellulose synthase catalytic chain [imported] - *Nostoc* sp. (strain PCC 7120)  
C;Species: *Nostoc* sp. PCC 7120  
A;Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C;Accession: AF2275  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.;  
Iriguchi, M.; Ishikawa, A.; Kawashima, K.; Kimura, T.; Kishida, Y.; Kohara, M.;  
Matsumoto, M.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.;  
Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing  
Cyanobacterium *Anabaena* sp. strain PCC 7120.  
A;Reference number: AB1807; MUID:21595285; PMID:11759840  
A;Accession: AF2275  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-693 <KUR>  
A;Cross-references: GB:BA000019; PIDN:BA075456.1; PID:g17132891; GSPDB:GN00179  
A;Experimental source: strain PCC 7120

C;Genetics:  
A;Gene: alr3757

Query Match 70.7%; Score 29; DB 2; Length 693;  
Best Local Similarity 62.5%; Pred. No. 4.1e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
| :||:|  
Db 420 LQYFADDF 427

RESULT 86

JH0595

potassium channel protein cdrK - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 20-Aug-1999

C;Accession: JH0595

R;Hwang, P.M.; Glatt, C.E.; Bredt, D.S.; Yellen, G.; Snyder, S.H.

Neuron 8, 473-481, 1992

A;Title: A novel K<sup>+</sup> channel with unique localizations in mammalian brain:  
molecular cloning and characterization.

A;Reference number: JH0595; MUID:92198655; PMID:1550672

A;Accession: JH0595

A;Molecule type: mRNA

A;Residues: 1-802 <HWA>

A;Cross-references: GB:M77482; NID:g203395; PIDN:AAA40905.1; PID:g203396

A;Experimental source: circumvallate papillae

C;Superfamily: potassium channel protein drk1

C;Keywords: glycoprotein; phosphoprotein; transmembrane protein

F;191-212/Domain: transmembrane #status predicted <TM1>

F;233-254/Domain: transmembrane #status predicted <TM2>

F;265-286/Domain: transmembrane #status predicted <TM3>

F;299-320/Domain: transmembrane #status predicted <TM4>

F;335-356/Domain: transmembrane #status predicted <TM5>

F;397-418/Domain: transmembrane #status predicted <TM6>

F;187,287,419,446/Binding site: carbohydrate (Asn) (covalent) #status predicted

F;448,500/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 70.7%; Score 29; DB 2; Length 802;  
Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAE 6  
| | | | |  
Db 352 LVFFAE 357

RESULT 87

F75294

DNA modification methyltransferase-related protein - Deinococcus radiodurans  
(strain R1)

C;Species: Deinococcus radiodurans

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000

C;Accession: F75294

R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.;

Dodson, R.J.; Haft, D.H.; Gwinn, M.L.; Nelson, W.C.; Richardson, D.L.; Moffat,

K.S.; Qin, H.; Jiang, L.; Pamphile, W.; Crosby, M.; Shen, M.; Vamathevan, J.J.;  
Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Makarova, K.S.; Aravind, L.;  
Daly, M.J.; Minton, K.W.; Fleischmann, R.D.; Ketchum, K.A.; Nelson, K.E.;  
Salzberg, S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999

A;Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans*  
R1.

A;Reference number: A75250; MUID:20036896; PMID:10567266

A;Accession: F75294

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-840 <WHI>

A;Cross-references: GB:AE002059; GB:AE000513; NID:g6460070; PIDN:AAF11813.1;

PID:g6460073; TIGR:DR2267; GSPDB:GN00077

A;Experimental source: strain R1

C;Genetics:

A;Gene: DR2267

A;Map position: 1

Query Match 70.7%; Score 29; DB 2; Length 840;  
Best Local Similarity 62.5%; Pred. No. 4.9e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
: |||||  
Db 195 MAMFAEDF 202

#### RESULT 88

CHRTD1

potassium channel protein drk1 - rat

C;Species: *Rattus norvegicus* (Norway rat)

C;Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 22-Jun-1999

C;Accession: S05448; A44838

R;Frech, G.C.; VanDongen, A.M.J.; Schuster, G.; Brown, A.M.; Joho, R.H.

Nature 340, 642-645, 1989

A;Title: A novel potassium channel with delayed rectifier properties isolated  
from rat brain by expression cloning.

A;Reference number: S05448; MUID:89365157; PMID:2770868

A;Accession: S05448

A;Molecule type: mRNA

A;Residues: 1-853 <FRE>

A;Cross-references: EMBL:X16476; NID:g57785; PIDN:CAA34497.1; PID:g57786

A;Note: it is uncertain whether Met-1 or Met-17 is the initiator

R;Drewe, J.A.; Verma, S.; Frech, G.; Joho, R.H.

J. Neurosci. 12, 538-548, 1992

A;Title: Distinct spatial and temporal expression patterns of K<sup>+</sup> channel mRNAs  
from different subfamilies.

A;Reference number: A44838; MUID:92156897; PMID:1740690

A;Accession: A44838

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 'MPAG', 1-571 <DRE>

A;Cross-references: GB:M81783; NID:g205038

A;Experimental source: brain

A;Note: sequence extracted from NCBI backbone (NCBIP:81768)

C;Genetics:

A;Gene: drk1  
C;Superfamily: potassium channel protein drk1  
C;Keywords: glycoprotein; ion channel; phosphoprotein; potassium channel;  
transmembrane protein  
F;1-182/Domain: intracellular #status predicted <INT1>  
F;183-204/Domain: transmembrane #status predicted <TM1>  
F;225-245/Domain: transmembrane #status predicted <TM2>  
F;256-276/Domain: transmembrane #status predicted <TM3>  
F;291-312/Domain: transmembrane #status predicted <TM4>  
F;327-348/Domain: transmembrane #status predicted <TM5>  
F;389-410/Domain: transmembrane #status predicted <TM6>  
F;411-853/Domain: intracellular #status predicted <INT2>  
F;279/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 70.7%; Score 29; DB 1; Length 853;  
Best Local Similarity 100.0%; Pred. No. 5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVFFAE 6  
| | | | |  
Db 344 LVFFAE 349

RESULT 89

I56529

potassium channel protein - mouse

C;Species: Mus musculus (house mouse)

C;Date: 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 20-Aug-1999

C;Accession: I56529

R;Pak, M.D.; Covarrubias, M.; Ratcliffe, A.; Salkoff, L.  
J. Neurosci. 11, 869-880, 1991

A;Title: A mouse brain homolog of the Drosophila Shab K+ channel with conserved  
delayed-rectifier properties.

A;Reference number: I56529; MUID:91162315; PMID:2002364

A;Accession: I56529

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-857 <RES>

A;Cross-references: GB:M64228; NID:g200975; PIDN:AAA40112.1; PID:g200976

C;Genetics:

A;Gene: Shab

C;Superfamily: potassium channel protein drk1

Query Match 70.7%; Score 29; DB 2; Length 857;  
Best Local Similarity 100.0%; Pred. No. 5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVFFAE 6  
| | | | |  
Db 348 LVFFAE 353

RESULT 90

S31761

potassium channel protein DRK1 - human

C;Species: Homo sapiens (man)

C;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 20-Aug-1999

C;Accession: S31761  
 R;Albrecht, B.; Lorra, C.; Stocker, K.; Pongs, O.  
 submitted to the EMBL Data Library, September 1992  
 A;Description: Cloning, expression and chromosomal localization of the delayed  
 rectifier type K<sup>+</sup> channel human 2.1 (h-DRK1) gene.  
 A;Reference number: S31761  
 A;Accession: S31761  
 A;Molecule type: DNA  
 A;Residues: 1-858 <ALB>  
 A;Cross-references: EMBL:X68302; NID:g30892; PIDN:CAA48374.1; PID:g30893  
 C;Genetics:  
 A;Gene: GDB:KCNB1; KV2.1; DRK1  
 A;Cross-references: GDB:l28081; OMIM:600397  
 A;Map position: 20q13.2-20q13.2  
 C;Superfamily: potassium channel protein drk1

Query Match 70.7%; Score 29; DB 2; Length 858;  
 Best Local Similarity 100.0%; Pred. No. 5.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAE 6  
 |||||  
 Db 348 LVFFAE 353

# RESULT 91

A69329

probable isoleucine-tRNA ligase (EC 6.1.1.5) - *Archaeoglobus fulgidus*

C;Species: *Archaeoglobus fulgidus*

C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 03-Jun-2002

C;Accession: A69329

R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum,  
 K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.;  
 Kerlavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush,  
 J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny,  
 K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger,  
 J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.;  
 McDonald, L.

Nature 390, 364-370, 1997

A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.;  
 Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;  
 Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing  
 archaeon *Archaeoglobus fulgidus*.

A;Reference number: A69250; MUID:98049343; PMID:9389475

A;Accession: A69329

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-1018 <KLE>

A;Cross-references: GB:AE001061; GB:AE000782; NID:g2689384; PIDN:AAB90608.1;  
 PID:g2649988; TIGR:AF0633

C;Superfamily: isoleucine-tRNA ligase

C;Keywords: ligase

Query Match 70.7%; Score 29; DB 2; Length 1018;  
 Best Local Similarity 83.3%; Pred. No. 6e+02;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy            3 FFAEDF 8  
              || |||  
Db            700 FFVEDE 705

RESULT 92

S72163

methyl-accepting chemotaxis protein mcpA - Rhizobium leguminosarum bv. viciae plasmid

N;Alternate names: chemoreceptor protein

C;Species: Rhizobium leguminosarum bv. viciae

C;Date: 04-Dec-1997 #sequence\_revision 12-Dec-1997 #text\_change 29-Sep-1999

C;Accession: S72163

R;Brito, B.; Palacios, J.M.; Ruiz-Argueeso, T.; Imperial, J.

Biochim. Biophys. Acta 1308, 7-11, 1996

A;Title: Identification of a gene for a chemoreceptor of the methyl-accepting type in the symbiotic plasmid of Rhizobium leguminosarum bv. viciae UPM791.

A;Reference number: S72162; MUID:96328256; PMID:8765742

A;Accession: S72163

A;Molecule type: DNA

A;Residues: 1-639 <BRI>

A;Cross-references: EMBL:U23040; NID:g780654; PIDN:AAC44311.1; PID:g780656

A;Experimental source: strain UPM791

C;Genetics:

A;Gene: mcpA

A;Genome: plasmid

A;Start codon: GTG

C;Superfamily: methyl-accepting chemotaxis protein mcpA

C;Keywords: chemotaxis; sensory transduction; transmembrane protein

Query Match            69.5%;    Score 28.5;    DB 2;    Length 639;  
Best Local Similarity    61.5%;    Pred. No. 4.8e+02;  
Matches        8;    Conservative        0;    Mismatches        0;    Indels        5;    Gaps        1;

Qy            1 LVF-----FAEDF 8  
              |||        |||||  
Db            315 LVFSLDDKFAEDF 327

RESULT 93

MNIHB2

nonstructural protein NS2 - bovine coronavirus

C;Species: bovine coronavirus

C;Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 16-Jul-1999

C;Accession: S08408; C46346

R;Woloszyn, N.; Boireau, P.; Laporte, J.

Nucleic Acids Res. 18, 1303, 1990

A;Title: Nucleotide sequence of the bovine enteric coronavirus BECV F15 mRNA 5 and mRNA 6 unique regions.

A;Reference number: S08408; MUID:90206809; PMID:2320429

A;Accession: S08408

A;Molecule type: mRNA

A;Residues: 1-109 <WOL>

A;Cross-references: EMBL:X51347; NID:g58684; PIDN:CAA35740.1; PID:g58685

A;Experimental source: strain F15

A;Note: the source is designated as bovine enteric coronavirus

R;Abraham, S.; Kienzle, T.E.; Lapps, W.E.; Brian, D.A.

Virology 177, 488-495, 1990

A;Title: Sequence and expression analysis of potential nonstructural proteins of 4.9, 4.8, 12.7, and 9.5 kDa encoded between the spike and membrane protein genes of the bovine coronavirus.

A;Reference number: A46346; MUID:90320120; PMID:2142556

A;Accession: C46346

A;Molecule type: genomic RNA

A;Residues: 1-109 <ABR>

A;Cross-references: GB:M31054; NID:g323363; PIDN:AAA42913.1; PID:g323367

A;Experimental source: strain Mebus

C;Superfamily: bovine coronavirus nonstructural protein NS2

C;Keywords: nonstructural protein

Query Match 68.3%; Score 28; DB 1; Length 109;  
Best Local Similarity 75.0%; Pred. No. 98;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
||| | ||  
Db 90 LVFLAVDF 97

#### RESULT 94

A44275

nonstructural protein NS2 - human coronavirus (strain OC43)

C;Species: human coronavirus

C;Date: 17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change 16-Jul-1999

C;Accession: A44275

R;Mounir, S.; Talbot, P.J.

Virology 192, 355-360, 1993

A;Title: Human coronavirus OC43 RNA 4 lacks two open reading frames located downstream of the S gene of bovine coronavirus.

A;Reference number: A44275; MUID:93297129; PMID:8517026

A;Accession: A44275

A;Molecule type: mRNA

A;Residues: 1-109 <MOU>

A;Cross-references: GB:M99576; NID:g329567; PIDN:AAA02569.1; PID:g329568

C;Superfamily: bovine coronavirus nonstructural protein NS2

C;Keywords: nonstructural protein

Query Match 68.3%; Score 28; DB 1; Length 109;  
Best Local Similarity 75.0%; Pred. No. 98;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
||| | ||  
Db 90 LVFLAVDF 97

#### RESULT 95

S58182

nonstructural protein, 12.8K - porcine hemagglutinating encephalomyelitis virus (strain 67N and strain NT9)

C;Species: porcine hemagglutinating encephalomyelitis virus

A;Variety: strain 67N; strain NT9

C;Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 20-Sep-1999



C;Accession: S58182; S58184  
R;Vieler, E.; Schlapp, T.; Herbst, W.  
submitted to the EMBL Data Library, July 1995  
A;Reference number: S58182  
A;Accession: S58182  
A;Molecule type: DNA  
A;Residues: 1-109 <VIE>  
A;Cross-references: EMBL:X89861; NID:g927492; PIDN:CAA61957.1; PID:g927493  
A;Experimental source: strain 67N  
A;Accession: S58184  
A;Molecule type: DNA  
A;Residues: 1-109 <VIW>  
A;Cross-references: EMBL:X89863; NID:g927495; PIDN:CAA61961.1; PID:g927496  
A;Experimental source: strain NT9  
C;Superfamily: bovine coronavirus nonstructural protein NS2  
C;Keywords: nonstructural protein

Query Match 68.3%; Score 28; DB 2; Length 109;  
Best Local Similarity 75.0%; Pred. No. 98;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
||| | ||  
Db 90 LVFLAVDF 97

RESULT 96

S58186  
nonstructural protein, 12.8K - porcine hemagglutinating encephalomyelitis virus  
(strain VW572)  
C;Species: porcine hemagglutinating encephalomyelitis virus  
A;Variety: strain VW572  
C;Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 20-Sep-1999  
C;Accession: S58186  
R;Vieler, E.; Schlapp, T.; Herbst, W.  
submitted to the EMBL Data Library, July 1995  
A;Reference number: S58182  
A;Accession: S58186  
A;Molecule type: DNA  
A;Residues: 1-109 <VIE>  
A;Cross-references: EMBL:X89862; NID:g927498; PIDN:CAA61959.1; PID:g927499  
A;Experimental source: strain VW572  
C;Superfamily: bovine coronavirus nonstructural protein NS2

Query Match 68.3%; Score 28; DB 2; Length 109;  
Best Local Similarity 75.0%; Pred. No. 98;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
||| | ||  
Db 90 LVFLAVDF 97

RESULT 97

R5HSS6  
ribosomal protein HS6 [validated] - Haloarcula marismortui  
C;Species: Haloarcula marismortui

C;Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 21-Jul-2000  
C;Accession: S00182  
R;Kimura, J.; Arndt, E.; Kimura, M.  
FEBS Lett. 224, 65-70, 1987  
A;Title: Primary structures of three highly acidic ribosomal proteins S6, S12 and S15 from the archaeobacterium Halobacterium marismortui.  
A;Reference number: S00182; MUID:88055606; PMID:3315748  
A;Accession: S00182  
A;Molecule type: protein  
A;Residues: 1-116 <KIM>  
A;Note: the source is designated as Halobacterium marismortui  
A;Note: the protein is designated as ribosomal protein S6  
A;Note: the protein was extracted from the 30S small ribosomal subunit  
C;Superfamily: rat ribosomal protein L7a  
C;Keywords: protein biosynthesis; ribosome

Query Match 68.3%; Score 28; DB 1; Length 116;  
Best Local Similarity 85.7%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
||| |||  
Db 47 LVFVAED 53

RESULT 98

H84375

Holliday junction resolvase [imported] - Halobacterium sp. NRC-1

C;Species: Halobacterium sp. NRC-1

C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 12-Jun-2003

C;Accession: H84375

R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.R.; Baliga, N.; Thorsson, V.; Sbrogna, J.; Swartzell, S.; Weir, D.; Hall, J.; Dahl, T.A.; Welte, R.; Goo, Y.A.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablonski, P.E.; Krebs, M.P.; Angevine, C.M.; Dale, H.; Isenbarger, T.A.; Peck, R.F.; Pohlschrod, M.; Spudich, J.L.; Jung, K.H.; Alam, M.; Freitas, T.

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Liang, P.; Riley, M.; Hood, L.; DasSarma, S.

A;Title: Genome sequence of Halobacterium species NRC-1.

A;Reference number: A84160; MUID:20504483; PMID:11016950

A;Accession: H84375

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-141 <STO>

A;Cross-references: GB:AE004437; NID:g10581665; PIDN:AAG20372.1; GSPDB:GN00138

C;Genetics:

A;Gene: hjr

C;Superfamily: uncharacterized conserved protein MJ0497

Query Match 68.3%; Score 28; DB 2; Length 141;  
Best Local Similarity 62.5%; Pred. No. 1.3e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
||:|| :|

## RESULT 99

C83475

hypothetical protein PA1354 [imported] - *Pseudomonas aeruginosa* (strain PA01)C;Species: *Pseudomonas aeruginosa*

C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C;Accession: C83475

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Brinkman, F.S.L.; Hufnagle, W.O.; Kowalik, D.J.; Lagrou, M.; Garber, R.L.; Goltry, L.; Tolentino, E.; Westbrook-Wadman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, R.M.; Smith, K.A.; Spencer, D.H.; Wong, G.K.S.; Wu, Z.; Paulsen, I.T.; Reizer, J.; Saier, M.H.; Hancock, R.E.W.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen.

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: C83475

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-141 &lt;STO&gt;

A;Cross-references: GB:AE004565; GB:AE004091; NID:g9947294; PIDN:AAG04743.1;

GSPDB:GN00131; PASP:PA1354

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA1354

Query Match 68.3%; Score 28; DB 2; Length 141;  
Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8

|| ||||

Db 112 VFFAEDF 118

## RESULT 100

S76052

hypothetical protein - *Synechocystis* sp. (strain PCC 6803)C;Species: *Synechocystis* sp.

A;Variety: PCC 6803

C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000

C;Accession: S76052

R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirosawa, M.; Sugiura, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 3, 109-136, 1996

A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.

A;Reference number: S74322; MUID:97061201; PMID:8905231

A;Accession: S76052

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-157 <KAN>

A;Cross-references: EMBL:D63999; GB:AB001339; NID:g1001396; PIDN:BAA10030.1;  
PID:g1001408

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June  
1996

C;Superfamily: protein-tyrosine-phosphatase, low molecular weight

Query Match 68.3%; Score 28; DB 2; Length 157;

Best Local Similarity 83.3%; Pred. No. 1.4e+02;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 FFAEDF 8

|| |||

Db 72 FFPEDF 77

Search completed: February 28, 2004, 08:56:50

Job time : 41 secs

OM protein - protein search, using sw model

Run on: February 28, 2004, 07:06:39 ; Search time 21.5 Seconds  
 (without alignments)  
 19.375 Million cell updates/sec

Title: US-09-668-314C-84  
 Perfect score: 41  
 Sequence: 1 LVFFAEDF 8

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 1000 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

# SUMMARIES

Result		%					
No.	Score	Query	Match	Length	DB	ID	Description
1	35	85.4	57	1	A4	URSMA	Q29149 ursus marit
2	35	85.4	58	1	A4	CANFA	Q28280 canis famil
3	35	85.4	58	1	A4	RABIT	Q28748 oryctolagus
4	35	85.4	58	1	A4	SHEEP	Q28757 ovis aries
5	35	85.4	59	1	A4	BOVIN	Q28053 bos taurus
6	35	85.4	751	1	A4	SAISC	Q95241 s amyloid b
7	35	85.4	770	1	A4	CAVPO	Q60495 c amyloid b
8	35	85.4	770	1	A4	HUMAN	P05067 h amyloid b
9	35	85.4	770	1	A4	MACFA	P53601 m amyloid b
10	35	85.4	770	1	A4	MOUSE	P12023 m amyloid b
11	35	85.4	770	1	A4	PIG	P79307 s amyloid b
12	35	85.4	770	1	A4	RAT	P08592 r amyloid b
13	35	85.4	780	1	A4	TETFL	O73683 tetraodon f
14	32	78.0	89	1	PE23	SHEEP	Q28550 ovis aries
15	32	78.0	737	1	A4	FUGRU	O93279 fugu rubrip
16	31	75.6	224	1	Y691	CHLTR	O84697 chlamydia t
17	31	75.6	281	1	UPK	CORST	Q9fb58 corynebacte

18	31	75.6	301	1	YWBI_BACSU	P39592	bacillus su
19	31	75.6	580	1	MM14_PIG	Q9xt90	sus scrofa
20	31	75.6	582	1	MM14_HUMAN	P50281	homo sapien
21	31	75.6	582	1	MM14_RABIT	Q95220	oryctolagus
22	31	75.6	622	1	YRT1_CAEEL	Q10044	caenorhabdi
23	31	75.6	956	1	MTN2_HUMAN	O00339	homo sapien
24	31	75.6	956	1	MTN2_MOUSE	O08746	mus musculu
25	31	75.6	1932	1	FAB1_SCHPO	O59722	schizosacch
26	31	75.6	2196	1	MOR2_SCHPO	Q9hdv6	schizosacch
27	30	73.2	224	1	Y681_CHLPN	Q9z7m3	chlamydia p
28	30	73.2	473	1	SYE_WIGBR	Q8d375	wiggleswort
29	30	73.2	529	1	YQP4_CAEEL	Q09531	caenorhabdi
30	30	73.2	570	1	GRAU_DROME	Q9u405	drosophila
31	30	73.2	641	1	LICR_BACSU	P46321	bacillus su
32	30	73.2	1006	1	BGAL_LACDE	P20043	lactobacill
33	30	73.2	1516	1	UGG2_HUMAN	Q9nyul	homo sapien
34	30	73.2	1888	1	CA1E_CHICK	P32018	gallus gall
35	29	70.7	251	1	Y126_PYRAB	Q9v2e8	pyrococcus
36	29	70.7	310	1	NU1M_DALCH	O63623	dalbulus ch
37	29	70.7	321	1	Y189_RICPR	Q9zdx5	rickettsia
38	29	70.7	357	1	HST2_YEAST	P53686	saccharomyc
39	29	70.7	380	1	HYD2_BRAJA	P31904	bradyrhizob
40	29	70.7	383	1	O94B_DROME	Q9vcs8	drosophila
41	29	70.7	385	1	HYPD_RHILV	P40598	rhizobium l
42	29	70.7	420	1	SYH_MYCPU	Q98qm8	mycoplasma
43	29	70.7	438	1	CLN3_CANFA	Q29611	canis famil
44	29	70.7	438	1	CLN3_MOUSE	Q61124	mus musculu
45	29	70.7	526	1	CH62_CHLPN	Q9z7c9	chlamydia p
46	29	70.7	663	1	UVRB_STRP3	Q8k743	streptococc
47	29	70.7	663	1	UVRB_STRP8	Q8p0j7	streptococc
48	29	70.7	663	1	UVRB_STRPY	Q99za5	streptococc
49	29	70.7	678	1	PTR3_YEAST	P43606	saccharomyc
50	29	70.7	809	1	KCB2_CANFA	Q95167	canis famil
51	29	70.7	857	1	KCB1_MOUSE	Q03717	mus musculu
52	29	70.7	857	1	KCB1_RAT	P15387	rattus norv
53	29	70.7	858	1	KCB1_HUMAN	Q14721	homo sapien
54	29	70.7	858	1	KCB1_PIG	O18868	sus scrofa
55	29	70.7	858	1	KCB1_RABIT	Q9mzl9	oryctolagus
56	29	70.7	907	1	KCB2_RAT	Q63099	rattus norv
57	29	70.7	911	1	KCB2_HUMAN	Q92953	homo sapien
58	29	70.7	911	1	KCB2_RABIT	Q95111	oryctolagus
59	29	70.7	1018	1	SYI_ARCFU	O29622	archaeoglob
60	28	68.3	109	1	VNS2_CVBM	P15774	bovine coro
61	28	68.3	109	1	VNS2_CVHOC	Q04853	human coron
62	28	68.3	119	1	RL7A_HALMA	P12743	haloarcula
63	28	68.3	157	1	Y328_SYNY3	Q55535	synechocyst
64	28	68.3	173	1	DESS_MYXXA	P02966	myxococcus
65	28	68.3	175	1	DEST_MYXXA	P02967	myxococcus
66	28	68.3	189	1	CME2_BACSU	P32393	bacillus su
67	28	68.3	196	1	BCB1_ARATH	Q07488	arabidopsis
68	28	68.3	214	1	UT11_ORYSA	Q8slzl	oryza sativ
69	28	68.3	246	1	Y928_THEMA	P56728	thermotoga
70	28	68.3	253	1	Y128_PYRHO	O57868	pyrococcus
71	28	68.3	275	1	APAH_ACTAC	O52655	actinobacil
72	28	68.3	293	1	Y347_HELPJ	Q9zmal	helicobacte
73	28	68.3	306	1	YC64_AQUAE	O67303	aquifex aeo
74	28	68.3	357	1	ALR2_PSEAE	Q9htq2	pseudomonas

75	28	68.3	372	1	YJGP_HAEIN	P45333	haemophilus
76	28	68.3	388	1	RAL1_ONCVO	P11012	onchocerca
77	28	68.3	392	1	PAN3_HUMAN	Q96qz0	homo sapien
78	28	68.3	401	1	ASSY_AQUAE	O67213	aquifex aeo
79	28	68.3	411	1	NQRF_HAEIN	O05012	haemophilus
80	28	68.3	421	1	CRT2_BOVIN	P42918	bos taurus
81	28	68.3	433	1	E1BL_ADEM1	P12536	mouse adeno
82	28	68.3	458	1	NTPB_ENTHR	Q08637	enterococcu
83	28	68.3	495	1	CRTD_RHOSH	Q01671	rhodobacter
84	28	68.3	533	1	SYK_METMP	O30522	methanococc
85	28	68.3	536	1	CH62_CHLCV	P59698	chlamydophi
86	28	68.3	538	1	COX1_RHILE	Q08855	rhizobium l
87	28	68.3	551	1	YABN_ECOLI	P33595	escherichia
88	28	68.3	575	1	APB3_HUMAN	O96018	homo sapien
89	28	68.3	576	1	DFA1_ANASP	Q8ynw5	anabaena sp
90	28	68.3	585	1	NEP2_THEVU	Q08751	thermoactin
91	28	68.3	697	1	CRAC_DICDI	P35401	dictyosteli
92	28	68.3	705	1	YKW2_YEAST	P35995	saccharomyc
93	28	68.3	728	1	VIV_ORYSA	P37398	oryza sativ
94	28	68.3	1056	1	YNN2_YEAST	P53914	saccharomyc
95	28	68.3	1353	1	CYA9_HUMAN	O60503	homo sapien
96	28	68.3	1353	1	CYA9_MOUSE	P51830	mus musculu
97	28	68.3	1894	1	BGS2_SCHPO	O13967	schizosacch
98	28	68.3	3137	1	CA36_CHICK	P15989	gallus gall
99	27	65.9	78	1	MIH_CANPG	P55846	cancer pagu
100	27	65.9	113	1	MIH_CALSI	P55321	callinectes
101	27	65.9	113	1	MIH_CANMG	O61389	cancer magi
102	27	65.9	113	1	MIH_CARMA	Q27225	carcinus ma
103	27	65.9	113	1	MIH_CHAFE	O96605	charybdis f
104	27	65.9	151	1	HS18_CLOAB	Q03928	clostridium
105	27	65.9	153	1	FLAP_HORSE	P30353	equus cabal
106	27	65.9	153	1	FLAP_PIG	P30356	sus scrofa
107	27	65.9	196	1	HIR5_HUMAN	Q9ums0	homo sapien
108	27	65.9	199	1	HIR5_MOUSE	Q9qz23	mus musculu
109	27	65.9	204	1	PSB3_SCHPO	Q9y7t8	schizosacch
110	27	65.9	204	1	YUFK_BACSU	O05249	bacillus su
111	27	65.9	217	1	MSA1_CAUCR	Q9a9i6	caulobacter
112	27	65.9	217	1	MSA1_RHIME	Q92sy7	rhizobium m
113	27	65.9	221	1	NU6M_PODAN	P15959	podospora a
114	27	65.9	246	1	PHOS_RAT	P20942	rattus norv
115	27	65.9	249	1	RNP1_YEAST	P32385	saccharomyc
116	27	65.9	258	1	TATC_ECOLI	P27857	escherichia
117	27	65.9	265	1	MUI2_BACSU	O05412	bacillus su
118	27	65.9	268	1	DNLI_HAEIN	P44121	haemophilus
119	27	65.9	271	1	YDJO_ECOLI	P76210	escherichia
120	27	65.9	273	1	LB20_ARATH	Q9srv3	arabidopsis
121	27	65.9	282	1	UPK_CHLTE	Q8kfj7	chlorobium
122	27	65.9	286	1	DMA_HAEIN	P44431	haemophilus
123	27	65.9	293	1	UPK_CORGL	Q8nqc3	corynebacte
124	27	65.9	294	1	YGXA_BACSU	Q04385	bacillus su
125	27	65.9	303	1	NQRF_ALTHA	Q9lcj4	alteromonas
126	27	65.9	303	1	NQRF_ALTMA	Q9lcj3	alteromonas
127	27	65.9	303	1	NQRF_SHEHA	Q9lci8	shewanella
128	27	65.9	303	1	NQRF_VIBMA	Q9lci9	vibrio mari
129	27	65.9	306	1	YC73_AQUAE	O67310	aquifex aeo
130	27	65.9	327	1	YFEI_CAEL	P83502	caenorhabdi
131	27	65.9	336	1	MTP2_PROVU	P11409	proteus vul

132	27	65.9	346	1	YR24_CAEEL	Q09341	caenorhabdi
133	27	65.9	362	1	VU8_HSV7J	P52523	human herpe
134	27	65.9	365	1	LMC1_HUMAN	Q9nzu5	homo sapien
135	27	65.9	378	1	MRG_HUMAN	P35410	homo sapien
136	27	65.9	380	1	TGT_STRP3	Q8p2s1	streptococc
137	27	65.9	380	1	TGT_STRPN	Q97nh1	streptococc
138	27	65.9	380	1	TGT_STRPY	Q9all6	streptococc
139	27	65.9	404	1	ASSY_LISMO	Q8y5h2	listeria mo
140	27	65.9	405	1	NQRF_NEIMA	Q9jvq3	neisseria m
141	27	65.9	405	1	NQRF_NEIMB	Q9k0m8	neisseria m
142	27	65.9	407	1	NQRF_PASMU	Q9cla6	pasteurella
143	27	65.9	407	1	NQRF_VIBAL	Q56584	v na(+)-tra
144	27	65.9	407	1	NQRF_VIBHA	Q9rfv6	vibrio harv
145	27	65.9	407	1	NQRF_VIBPA	Q9lcj0	vibrio para
146	27	65.9	407	1	NQRF_VIBVU	Q8dbj1	vibrio vuln
147	27	65.9	408	1	NQRF_VIBCH	Q9x4q8	vibrio chol
148	27	65.9	442	1	GNT2_RAT	Q09326	r alpha-1,6
149	27	65.9	446	1	TEB_EUPCR	Q06184	euplotes cr
150	27	65.9	447	1	GNT2_HUMAN	Q10469	h alpha-1,6
151	27	65.9	460	1	YS41_CAEEL	Q09370	caenorhabdi
152	27	65.9	468	1	LIP1_CANAL	O94091	candida alb
153	27	65.9	468	1	UXAC_BACTN	Q8a9j2	bacteroides
154	27	65.9	470	1	RBL_SYNPF	Q44176	synechococc
155	27	65.9	493	1	SYK_ARCFU	O29052	archaeoglob
156	27	65.9	501	1	ABC1_YEAST	P27697	saccharomyc
157	27	65.9	501	1	AMPA_BUCBP	Q89ag2	buchnera ap
158	27	65.9	513	1	NDI1_YEAST	P32340	saccharomyc
159	27	65.9	543	1	C307_DROME	Q9vrm7	drosophila
160	27	65.9	560	1	TATR_NPVOP	P22114	orgyia pseu
161	27	65.9	588	1	Z429_HUMAN	Q86v71	homo sapien
162	27	65.9	610	1	CALG_HUMAN	O14967	homo sapien
163	27	65.9	631	1	PRKA_BACSU	P39134	bacillus su
164	27	65.9	672	1	NXF1_DROME	Q9ulh9	drosophila
165	27	65.9	685	1	CRPI_PERAM	Q25641	periplaneta
166	27	65.9	697	1	TALA_POVLY	P04008	lymphotropi
167	27	65.9	734	1	GLGB_NITEU	Q81zu6	nitrosomona
168	27	65.9	768	1	ITB8_RABIT	P26013	oryctolagus
169	27	65.9	769	1	ITB8_HUMAN	P26012	homo sapien
170	27	65.9	774	1	YQV5_CAEEL	Q09554	caenorhabdi
171	27	65.9	930	1	SYI_STRPN	Q9zhb3	streptococc
172	27	65.9	1071	1	TRI_THEAC	P96086	thermoplasm
173	27	65.9	1286	1	RPOC_MYCGA	P47716	mycoplasma
174	27	65.9	1713	1	LMA3_HUMAN	Q16787	homo sapien
175	27	65.9	1785	1	GLS3_YEAST	Q04952	saccharomyc
176	27	65.9	2164	1	CCAA_MOUSE	P97445	mus musculu
177	27	65.9	2167	1	BPH1_YEAST	P25356	saccharomyc
178	27	65.9	3333	1	LMA3_MOUSE	Q61789	mus musculu
179	27	65.9	3375	1	UN52_CAEEL	Q06561	caenorhabdi
180	27	65.9	3746	1	ACVS_PENCH	P19787	penicillium
181	27	65.9	3791	1	ACVT_PENCH	P26046	penicillium
182	26.5	64.6	484	1	ER60_SCHMA	P38658	schistosoma
183	26	63.4	90	1	HSB7_RAT	Q9quk5	rattus norv
184	26	63.4	115	1	YACC_ECOLI	P23838	escherichia
185	26	63.4	119	1	RL7A_ARCFU	O29494	archaeoglob
186	26	63.4	120	1	RL7A_HALN1	Q9hgh8	halobacteri
187	26	63.4	128	1	YMJ0_YEAST	Q04501	saccharomyc
188	26	63.4	153	1	FLAP_SHEEP	P30358	ovis aries



189	26	63.4	159	1	YM38_STRPN	P72565	streptococc
190	26	63.4	162	1	COAT_SHMV	P03581	sunh-hemp m
191	26	63.4	162	1	RRAA_HAEIN	P44738	haemophilus
192	26	63.4	165	1	RRAA_HAEDU	P59885	haemophilus
193	26	63.4	165	1	YL72_AQUAE	O67921	aquifex aeo
194	26	63.4	166	1	RRAA_PASMU	Q9clp9	pasteurella
195	26	63.4	200	1	COAE_MYCPN	P75400	mycoplasma
196	26	63.4	202	1	B3G1_MOUSE	Q9cw73	m galactosy
197	26	63.4	210	1	END3_BUCAI	P57219	buchnera ap
198	26	63.4	218	1	TPMT_VIBCH	Q9ksn0	vibrio chol
199	26	63.4	222	1	GL18_ARATH	Q9lea7	arabidopsis
200	26	63.4	225	1	YEA1_YEAST	P40006	saccharomyc
201	26	63.4	225	1	YJYG_ECOLI	P33999	escherichia
202	26	63.4	233	1	CYSA_BACCE	O31339	bacillus ce
203	26	63.4	239	1	FLIP_AQUAE	O67750	aquifex aeo
204	26	63.4	240	1	Y833_TREPA	O83805	treponema p
205	26	63.4	245	1	Y191_LISMO	Q8yae0	listeria mo
206	26	63.4	245	1	Y230_LISIN	Q92f76	listeria in
207	26	63.4	255	1	MTL4_HUMAN	Q8tcb7	homo sapien
208	26	63.4	255	1	Y439_PYRHO	O58186	pyrococcus
209	26	63.4	257	1	FCEA_HUMAN	P12319	homo sapien
210	26	63.4	267	1	Y012_PYRFU	Q8u4r1	pyrococcus
211	26	63.4	271	1	HMGL_TRYBR	P26586	trypanosoma
212	26	63.4	277	1	ICA2_MOUSE	P35330	mus musculu
213	26	63.4	280	1	GPI2_YEAST	P46961	saccharomyc
214	26	63.4	292	1	XTHA_PHAAN	Q41638	phaseolus a
215	26	63.4	293	1	XT31_ARATH	P93046	arabidopsis
216	26	63.4	293	1	XTH5_ARATH	Q9xiw1	arabidopsis
217	26	63.4	300	1	PRMA_LEPIN	Q8f6b7	leptospira
218	26	63.4	303	1	NQRF_PHOPO	Q9lcj1	photobacter
219	26	63.4	320	1	RIR2_HSVE4	P50644	equine herp
220	26	63.4	323	1	B3G2_HUMAN	Q9npz5	homo sapien
221	26	63.4	324	1	B3G2_MOUSE	P59270	mus musculu
222	26	63.4	324	1	B3G2_RAT	Q9z137	rattus norv
223	26	63.4	327	1	MALR_STRPN	Q08511	streptococc
224	26	63.4	327	1	RRPP_PIRYV	Q01769	piry virus.
225	26	63.4	331	1	HOLA_BUCAI	P57520	buchnera ap
226	26	63.4	334	1	ANM4_HUMAN	Q9nr22	homo sapien
227	26	63.4	334	1	B3G1_HUMAN	Q9p2w7	homo sapien
228	26	63.4	334	1	B3G1_RAT	O35789	r galactosy
229	26	63.4	335	1	B3G3_CRIGR	Q9wu47	cricketulus
230	26	63.4	335	1	B3G3_HUMAN	O94766	homo sapien
231	26	63.4	335	1	B3G3_MOUSE	P58158	mus musculu
232	26	63.4	352	1	PAP2_CARPA	P14080	carica papa
233	26	63.4	353	1	ANM1_RAT	Q63009	rattus norv
234	26	63.4	357	1	CYSA_BACCR	Q8lgul	bacillus ce
235	26	63.4	361	1	ANM1_HUMAN	Q99873	homo sapien
236	26	63.4	363	1	GCST_ECOLI	P27248	escherichia
237	26	63.4	364	1	GCST_ECO57	Q8xd32	escherichia
238	26	63.4	364	1	GCST_ECOL6	Q8fe65	escherichia
239	26	63.4	364	1	GCST_SALTY	Q8xg67	salmonella
240	26	63.4	365	1	Y255_MYCGE	P47497	mycoplasma
241	26	63.4	371	1	ANM1_MOUSE	Q9jif0	mus musculu
242	26	63.4	371	1	RECF_BACHD	Q9rc99	bacillus ha
243	26	63.4	373	1	ACT_CHOCR	P53499	chondrus cr
244	26	63.4	373	1	VU3_HSV6U	Q01350	human herpe
245	26	63.4	376	1	ACT_ENTHI	P11426	entamoeba h

246	26	63.4	379	1	CYB_MARBA	Q9th43	marmota bai
247	26	63.4	379	1	CYB_MYONA	Q956z9	myotis natt
248	26	63.4	379	1	CYB_PROTA	Q35677	propithecus
249	26	63.4	382	1	LGT_MYCGE	P47332	mycoplasma
250	26	63.4	387	1	YM14_PARTE	P15615	paramecium
251	26	63.4	388	1	ALGJ_AZOVI	O52197	azotobacter
252	26	63.4	394	1	AV71_ACAVI	Q17107	acanthochei
253	26	63.4	398	1	ACTU_DROME	P45890	drosophila
254	26	63.4	398	1	DXR_BUCAI	P57329	buchnera ap
255	26	63.4	404	1	LA_BOVIN	P10881	bos taurus
256	26	63.4	405	1	CYB_RHORU	P23134	rhodospiril
257	26	63.4	407	1	NQRF_YERPE	Q8zbz5	yersinia pe
258	26	63.4	415	1	LA_MOUSE	P32067	mus musculu
259	26	63.4	415	1	LA_RAT	P38656	rattus norv
260	26	63.4	416	1	CRTC_NICPL	Q40401	nicotiana p
261	26	63.4	417	1	HS47_HUMAN	P29043	homo sapien
262	26	63.4	417	1	HS47_MOUSE	P19324	mus musculu
263	26	63.4	417	1	HS47_RAT	P29457	rattus norv
264	26	63.4	417	1	O65A_DROME	P82982	drosophila
265	26	63.4	418	1	SIH2_HUMAN	P50454	homo sapien
266	26	63.4	425	1	RAFB_ECOLI	P16552	escherichia
267	26	63.4	428	1	SYH_BUCAP	Q8k9p3	buchnera ap
268	26	63.4	432	1	OV71_ONCVO	P31732	onchocerca
269	26	63.4	438	1	CLN3_HUMAN	Q13286	homo sapien
270	26	63.4	449	1	CPSE_STRA3	Q04664	streptococc
271	26	63.4	449	1	CPSE_STRA5	Q9afi0	streptococc
272	26	63.4	459	1	RBL_CALSH	P48687	calyptrosph
273	26	63.4	459	1	RBL_CHRHI	P48692	chrysochrom
274	26	63.4	470	1	RBL_SYNY3	P54205	synechocyst
275	26	63.4	471	1	SYW_HUMAN	P23381	homo sapien
276	26	63.4	475	1	RBL_ASTLO	P26490	astasia lon
277	26	63.4	475	1	SYW_BOVIN	P17248	bos taurus
278	26	63.4	475	1	SYW_RABIT	P23612	oryctolagus
279	26	63.4	481	1	SYW_MOUSE	P32921	mus musculu
280	26	63.4	488	1	RBL_GUITH	P14957	guillardia
281	26	63.4	488	1	RBL_PLECA	Q08051	pleurochrys
282	26	63.4	500	1	PCP1_MOUSE	P20943	mus musculu
283	26	63.4	507	1	PDI_DATGL	Q9xf61	datisca glo
284	26	63.4	511	1	VGLG_VSVIG	P04883	vesicular s
285	26	63.4	511	1	VGLG_VSVO	P04884	vesicular s
286	26	63.4	511	1	VGLG_VSVSJ	P03522	vesicular s
287	26	63.4	512	1	CADC_ECOLI	P23890	escherichia
288	26	63.4	516	1	HS70_LEIMA	P14834	leishmania
289	26	63.4	525	1	BCHB_RHOCA	P26163	rhodobacter
290	26	63.4	534	1	BCHB_RHOSH	Q9z5d9	rhodobacter
291	26	63.4	534	1	Y255_MYCPN	P75422	mycoplasma
292	26	63.4	541	1	COX1_BRAJA	P31833	bradyrhizob
293	26	63.4	569	1	CH63_HELVI	P25420	heliobacter
294	26	63.4	595	1	SYD_BACHD	Q9kdgl	bacillus ha
295	26	63.4	602	1	SYD_RICCN	Q92j82	rickettsia
296	26	63.4	610	1	VG17_BPT4	P17312	bacterioph
297	26	63.4	618	1	MM24_MOUSE	Q9r0s2	mus musculu
298	26	63.4	618	1	MM24_RAT	Q99pw6	rattus norv
299	26	63.4	630	1	CREC_EMENI	Q9p4r5	emerella
300	26	63.4	645	1	MM24_HUMAN	Q9y5r2	homo sapien
301	26	63.4	646	1	UVRB_METTH	O26542	methanobact
302	26	63.4	652	1	HS70_LEIAM	Q07437	leishmania

303	26	63.4	653	1	HS70_LEIDO	P17804	leishmania
304	26	63.4	656	1	UVRB_CHLCV	Q821h8	chlamydophi
305	26	63.4	657	1	UVRB_CHLPN	Q9z7a5	chlamydia p
306	26	63.4	659	1	UVRB_CLOPE	Q46323	clostridium
307	26	63.4	660	1	UVRB_BACHD	Q9k6x9	bacillus ha
308	26	63.4	662	1	UVRB_STRPN	Q54986	streptococc
309	26	63.4	662	1	UVRB_STRR6	Q8dpk7	streptococc
310	26	63.4	662	1	UVRB_THETN	Q8r8m4	thermoanaer
311	26	63.4	663	1	UVRB_AQUAE	O67708	aquifex aeo
312	26	63.4	665	1	UVRB_THETH	Q56243	thermus the
313	26	63.4	670	1	UVRB_METAC	Q8tk33	methanosarc
314	26	63.4	676	1	UVRB_CHLMU	Q9pjf4	chlamydia m
315	26	63.4	692	1	UVRB_LACLA	Q9ci06	lactococcus
316	26	63.4	698	1	UVRB_MYCLE	P57991	mycobacteri
317	26	63.4	698	1	UVRB_MYCTU	O06150	mycobacteri
318	26	63.4	702	1	ARYA_MANSE	P14296	manduca sex
319	26	63.4	703	1	ARYB_MANSE	P14297	manduca sex
320	26	63.4	703	1	PURL_SULTO	Q970v6	sulfolobus
321	26	63.4	704	1	SSP2_BOMMO	P20613	bombyx mori
322	26	63.4	734	1	PSAB_EUGGR	P19431	euglena gra
323	26	63.4	749	1	JSB2_TRINI	Q06343	trichoplusi
324	26	63.4	785	1	PTA1_YEAST	Q01329	saccharomyc
325	26	63.4	790	1	SUV6_ARATH	Q8vz17	arabidopsis
326	26	63.4	834	1	RSG3_BOVIN	Q28013	bos taurus
327	26	63.4	834	1	RSG3_HUMAN	Q14644	homo sapien
328	26	63.4	834	1	RSG3_MOUSE	Q60790	mus musculu
329	26	63.4	849	1	RSG2_HUMAN	Q15283	homo sapien
330	26	63.4	862	1	LOX1_HORVU	P29114	hordeum vul
331	26	63.4	865	1	LAC9_KLULA	P08657	kluveromyc
332	26	63.4	879	1	YN65_YEAST	P42837	saccharomyc
333	26	63.4	901	1	R8L3_ARATH	Q9fjb5	arabidopsis
334	26	63.4	904	1	PPR1_YEAST	P07272	saccharomyc
335	26	63.4	908	1	R8L4_ARATH	Q9fjk8	arabidopsis
336	26	63.4	908	1	RPP8_ARATH	Q8w4j9	arabidopsis
337	26	63.4	910	1	RP8H_ARATH	P59584	arabidopsis
338	26	63.4	968	1	Y080_BUCAP	Q8ka32	buchnera ap
339	26	63.4	983	1	CAGE_HELPJ	Q9zlt4	helicobacte
340	26	63.4	983	1	CAGE_HELPY	Q48252	helicobacte
341	26	63.4	1042	1	T1RH_METJA	Q60295	methanococc
342	26	63.4	1075	1	Y124_METJA	Q57588	methanococc
343	26	63.4	1088	1	SEC8_SCHPO	O74562	schizosacch
344	26	63.4	1163	1	CQAA_BACTF	Q9x597	bacillus th
345	26	63.4	1358	1	MSH6_MOUSE	P54276	mus musculu
346	26	63.4	1360	1	MSH6_HUMAN	P52701	homo sapien
347	26	63.4	1447	1	SGS1_YEAST	P35187	saccharomyc
348	26	63.4	1523	1	DIP2_MOUSE	Q8bwt5	mus musculu
349	26	63.4	1556	1	Y934_HUMAN	Q9y2e4	homo sapien
350	26	63.4	1571	1	DIP2_HUMAN	Q14689	homo sapien
351	26	63.4	1587	1	SUR2_CAEEL	Q10669	caenorhabdi
352	26	63.4	1790	1	SEPA_EMENI	P78621	emericella
353	26	63.4	2096	1	BP28_DROME	Q9vm75	drosophila
354	26	63.4	2199	1	DPOE_SCHPO	P87154	schizosacch
355	26	63.4	2670	1	IP3T_RAT	Q63269	rattus norv
356	26	63.4	2671	1	IP3T_HUMAN	Q14573	homo sapien
357	26	63.4	2709	1	IP3R_BOVIN	Q9tu34	bos taurus
358	26	63.4	2749	1	IP3R_MOUSE	P11881	m inositol
359	26	63.4	2750	1	IP3R_RAT	P29994	rattus norv

360	26	63.4	2758	1	IP3R_HUMAN	Q14643	homo sapien
361	26	63.4	2834	1	IP3R_DROME	P29993	drosophila
362	26	63.4	3068	1	POLG_PEMVC	Q01500	p genome po
363	26	63.4	5327	1	MACF_MOUSE	Q9qxz0	mus musculu
364	26	63.4	5430	1	MACF_HUMAN	Q9upn3	homo sapien
365	26	63.4	5938	1	MAC4_HUMAN	Q96pk2	homo sapien
366	25	61.0	15	1	LMA2_LOCF1	P38497	locusta mig
367	25	61.0	109	1	YCCK_HAEIN	P45184	haemophilus
368	25	61.0	111	1	YG3U_YEAST	P53291	saccharomyc
369	25	61.0	112	1	YH01_ENTFA	Q834f4	enterococcu
370	25	61.0	117	1	YCX2_CHLRE	P05722	chlamydomon
371	25	61.0	131	1	Y624_MYCTU	P96914	mycobacteri
372	25	61.0	149	1	NDK_TREPA	O83974	treponema p
373	25	61.0	153	1	FLAP_MACMU	P30354	macaca mula
374	25	61.0	153	1	VJ01_VACCC	P21032	vaccinia vi
375	25	61.0	153	1	VJ01_VACCV	P07616	vaccinia vi
376	25	61.0	159	1	ISPF_ECOL6	Q8fej6	escherichia
377	25	61.0	159	1	ISPF_ECOLI	P36663	escherichia
378	25	61.0	159	1	ISPF_SALTI	Q8z472	salmonella
379	25	61.0	159	1	ISPF_SALTY	Q8zmf7	salmonella
380	25	61.0	159	1	VJ01_VARV	P33004	variola vir
381	25	61.0	160	1	OBP2_ANTPE	Q17075	antheraea p
382	25	61.0	161	1	FLAP_HUMAN	P20292	homo sapien
383	25	61.0	161	1	FLAP_RAT	P20291	rattus norv
384	25	61.0	163	1	YCBL_BACUN	P30906	bacteroides
385	25	61.0	166	1	YVAE_VACCC	P20514	vaccinia vi
386	25	61.0	167	1	OBP1_ANTPE	P87508	antheraea p
387	25	61.0	167	1	THIX_HAEIN	P43787	haemophilus
388	25	61.0	181	1	CS22_BACSU	P94497	bacillus su
389	25	61.0	182	1	PTHC_CLOBE	O32332	clostridium
390	25	61.0	190	1	RECR_CAMJE	Q9pn35	campylobact
391	25	61.0	198	1	COAE_MYCGE	P47506	mycoplasma
392	25	61.0	201	1	RL4_BUCAI	P57590	buchnera ap
393	25	61.0	218	1	MSRA_OCHAN	Q93s39	ochrobactru
394	25	61.0	228	1	UT11_ARATH	Q9m223	arabidopsis
395	25	61.0	235	1	YWFC_BACSU	P39639	bacillus su
396	25	61.0	236	1	ALDC_LACLA	P95676	lactococcus
397	25	61.0	236	1	ALDC_LACLC	P77880	lactococcus
398	25	61.0	236	1	Y264_METEX	Q8gek8	methylobact
399	25	61.0	244	1	YH38_HAEIN	P44302	haemophilus
400	25	61.0	246	1	YM95_CLOAB	Q97gs1	clostridium
401	25	61.0	251	1	YQ33_VIBPA	Q87lh9	vibrio para
402	25	61.0	252	1	YC85_VIBCH	Q9ksh1	vibrio chol
403	25	61.0	252	1	YE86_VIBVU	Q8dcd5	vibrio vuln
404	25	61.0	252	1	YS97_VIBVY	Q7mhh4	vibrio vuln
405	25	61.0	257	1	CT53_HUMAN	Q9hln1	homo sapien
406	25	61.0	257	1	PEBB_PROMA	Q9k4u5	prochloroco
407	25	61.0	257	1	PYRK_BACCL	P46536	bacillus ca
408	25	61.0	257	1	Y2G9_BACTN	Q8a102	bacteroides
409	25	61.0	260	1	ARGB_BACHD	Q9k8v4	bacillus ha
410	25	61.0	265	1	DAPB_WIGBR	Q8d3h6	wiggleswort
411	25	61.0	267	1	XT11_ARATH	Q9smp1	arabidopsis
412	25	61.0	270	1	DHMA_FLAS1	P22441	flavobacter
413	25	61.0	276	1	DAPF_VIBPA	Q87kj4	vibrio para
414	25	61.0	278	1	UPK1_STRCO	Q9fc36	streptomyce
415	25	61.0	280	1	PARB_CHLMU	Q9pln9	chlamydia m
416	25	61.0	283	1	POR1_YEAST	P04840	saccharomyc

417	25	61.0	290	1	F16P_HELPJ	Q9zj74	helicobacte
418	25	61.0	290	1	F16P_HELPY	O25936	helicobacte
419	25	61.0	298	1	ADT2_HUMAN	P05141	homo sapien
420	25	61.0	298	1	ADT2_MOUSE	P51881	mus musculu
421	25	61.0	298	1	ADT2_RAT	Q09073	rattus norv
422	25	61.0	299	1	VG12_HSVI1	Q00165	ictalurid h
423	25	61.0	300	1	ER25_SCHPO	Q9uu4	schizosacch
424	25	61.0	303	1	RIR2_PRVKA	P50645	pseudorabie
425	25	61.0	303	1	YN37_ARCFU	O27947	archaeoglob
426	25	61.0	306	1	B3GI_DROME	O97422	d galactosy
427	25	61.0	306	1	BXD1_HUMAN	Q9h7b2	homo sapien
428	25	61.0	306	1	BXD1_MOUSE	Q9jj80	mus musculu
429	25	61.0	321	1	DDL_RICCN	Q92it7	rickettsia
430	25	61.0	321	1	K6PF_AQUAE	O67605	aquifex aeo
431	25	61.0	326	1	RLUD_BORBU	P70870	borrelia bu
432	25	61.0	328	1	KDSD_ECO57	Q8x9j0	escherichia
433	25	61.0	328	1	KDSD_ECOL6	Q8fd73	escherichia
434	25	61.0	328	1	KDSD_ECOLI	P45395	escherichia
435	25	61.0	328	1	KDSD_SALTI	Q8z3g6	salmonella
436	25	61.0	328	1	KDSD_SALTY	Q8zls1	salmonella
437	25	61.0	328	1	KDSD_SHIFL	Q83jf4	shigella fl
438	25	61.0	337	1	RIR2_HSV23	P03174	herpes simp
439	25	61.0	341	1	ACOB_BACSU	O34591	bacillus su
440	25	61.0	341	1	Y665_METJA	Q58079	methanococc
441	25	61.0	350	1	LSS1_MOUSE	P27545	mus musculu
442	25	61.0	350	1	YCXA_EUGGR	P31561	euglena gra
443	25	61.0	354	1	GB01_DROME	P16378	drosophila
444	25	61.0	354	1	GB0_LOCFI	P38404	locusta mig
445	25	61.0	357	1	CYSA_BACHD	Q9k876	bacillus ha
446	25	61.0	357	1	YCT4_YEAST	P25625	saccharomyc
447	25	61.0	358	1	V044_FOWPV	Q9j5f9	fowlpox vir
448	25	61.0	359	1	AG2R_CHICK	P79785	gallus gall
449	25	61.0	359	1	AG2R_MELGA	P33396	meleagris g
450	25	61.0	361	1	RL4A_YEAST	P10664	saccharomyc
451	25	61.0	361	1	RL4B_YEAST	P49626	saccharomyc
452	25	61.0	362	1	Y593_CHLPN	Q9z7wl	chlamydia p
453	25	61.0	365	1	FXH1_HUMAN	O75593	homo sapien
454	25	61.0	370	1	GP85_HUMAN	Q9npdl	homo sapien
455	25	61.0	371	1	GP85_BRARE	Q9i919	brachydanio
456	25	61.0	371	1	PRIL_METAC	Q8tuf8	methanosarc
457	25	61.0	378	1	O33A_DROME	P81914	drosophila
458	25	61.0	379	1	YISP_BACCS	P29156	bacillus cl
459	25	61.0	380	1	CRT3_MOUSE	Q9d9q6	mus musculu
460	25	61.0	384	1	CRT3_HUMAN	Q96112	homo sapien
461	25	61.0	384	1	VU3_HSV7J	P52520	human herpe
462	25	61.0	385	1	SYW_PYRAB	Q9uy11	pyrococcus
463	25	61.0	386	1	SYW_PYRHO	O59584	pyrococcus
464	25	61.0	392	1	XYLR_ECOLI	P37390	escherichia
465	25	61.0	393	1	NIFS_BRAJA	P37030	bradyrhizob
466	25	61.0	395	1	CRTC_CAEEL	P27798	caenorhabdi
467	25	61.0	395	1	NIFS_AZOCH	P23120	azotobacter
468	25	61.0	400	1	NIFS_ACEDI	P57794	acetobacter
469	25	61.0	401	1	NIFS_AZOVI	P05341	azotobacter
470	25	61.0	417	1	PQIA_ECOLI	P43670	escherichia
471	25	61.0	417	1	YAG1_YEAST	P39713	saccharomyc
472	25	61.0	420	1	CRTC_MAIZE	Q9sp22	zea mays (m
473	25	61.0	420	1	GLYA_STRCO	O86565	streptomyce

474	25	61.0	424	1	CRTC_ORYSA	Q9sly8	oryza sativ
475	25	61.0	429	1	GLYA_METJA	Q58992	methanococc
476	25	61.0	432	1	PEPB_VIBPA	Q87s21	vibrio para
477	25	61.0	432	1	SURA_BUCAP	Q8ka01	buchnera ap
478	25	61.0	444	1	BCHN_CHLAU	Q9f6x6	chloroflexu
479	25	61.0	444	1	VU5_HSV6U	Q01352	human herpe
480	25	61.0	446	1	AP50_SCHPO	Q09718	schizosacch
481	25	61.0	448	1	SDHD_BACSU	P54555	bacillus su
482	25	61.0	448	1	YB00_METJA	Q58500	methanococc
483	25	61.0	453	1	BIOA_AQUAE	O66557	aquifex aeo
484	25	61.0	458	1	YF10_MYCPN	P75276	mycoplasma
485	25	61.0	464	1	BIOA_METJA	Q58696	methanococc
486	25	61.0	465	1	RBL_NEPAL	P28434	nepenthes a
487	25	61.0	466	1	SYN_YERPE	P58697	yersinia pe
488	25	61.0	470	1	RBL_PROHO	P27568	prochloroth
489	25	61.0	472	1	RBL_SYNP6	P00880	synechococc
490	25	61.0	473	1	CYSG_BUCAI	P57500	buchnera ap
491	25	61.0	473	1	VGLM_HSV11	P04288	herpes simp
492	25	61.0	478	1	RBL_NEUMU	P19163	neurachne m
493	25	61.0	478	1	RBL_NEUTE	P19164	neurachne t
494	25	61.0	480	1	Y066_METJA	Q60377	methanococc
495	25	61.0	483	1	SYE_DEIRA	Q9rx30	deinococcus
496	25	61.0	485	1	PUR1_VIGAC	P52419	vigna aconi
497	25	61.0	490	1	VATB_DROME	P31409	drosophila
498	25	61.0	491	1	VATB_CAEEL	Q19626	caenorhabdi
499	25	61.0	494	1	VATB_HELVI	P31410	heliothis v
500	25	61.0	494	1	VATB_MANSE	P31401	manduca sex
501	25	61.0	497	1	MEK1_YEAST	P24719	saccharomyc
502	25	61.0	497	1	RELB_XENLA	P51510	xenopus lae
503	25	61.0	499	1	CIK2_XENLA	P22739	xenopus lae
504	25	61.0	500	1	VATB_CYACA	P48413	cyanidium c
505	25	61.0	501	1	COBQ_BACHD	Q9kci0	bacillus ha
506	25	61.0	501	1	SCRB_PEDPE	P43471	pediococcus
507	25	61.0	506	1	YF26_SYNY3	P74360	synechocyst
508	25	61.0	509	1	HUTH_PSEPU	P21310	pseudomonas
509	25	61.0	511	1	VATB_CANTR	P22550	candida tro
510	25	61.0	512	1	FEN2_YEAST	P25621	saccharomyc
511	25	61.0	514	1	LIM_HALRO	Q25132	halocynthia
512	25	61.0	515	1	SIR2_CANAL	O59923	candida alb
513	25	61.0	517	1	VATB_YEAST	P16140	saccharomyc
514	25	61.0	522	1	G6PD_CAEEL	Q27464	caenorhabdi
515	25	61.0	531	1	C7E1_SORBI	O48958	sorghum bic
516	25	61.0	532	1	VAT_HUMAN	Q16572	homo sapien
517	25	61.0	536	1	MVIN_CHLMU	Q9pjb9	chlamydia m
518	25	61.0	536	1	MVIN_CHLTR	Q46378	chlamydia t
519	25	61.0	537	1	FUT6_ARATH	Q9xi80	arabidopsis
520	25	61.0	546	1	CHK2_MOUSE	Q9z265	mus musculu
521	25	61.0	554	1	CX1A_PARDE	P08305	paracoccus
522	25	61.0	558	1	CX1B_PARDE	P98002	paracoccus
523	25	61.0	574	1	G6PC_SPIOL	O24357	spinacia ol
524	25	61.0	574	1	HMD3_SOLTU	Q41438	solanum tub
525	25	61.0	576	1	GPD1_ARATH	Q43727	arabidopsis
526	25	61.0	580	1	ACEA_PINTA	Q43097	pinus taeda
527	25	61.0	580	1	SRK1_SCHPO	O94547	schizosacch
528	25	61.0	582	1	MM14_MOUSE	P53690	mus musculu
529	25	61.0	582	1	MM14_RAT	Q10739	rattus norv
530	25	61.0	591	1	CALX_MOUSE	P35564	mus musculu

531	25	61.0	591	1	CALX_RAT	P35565	rattus norv
532	25	61.0	592	1	CALX_HUMAN	P27824	homo sapien
533	25	61.0	592	1	EPHD_MYCTU	Q10402	mycobacteri
534	25	61.0	593	1	CALX_CANFA	P24643	canis famil
535	25	61.0	594	1	YKM1_SCHPO	Q9c0z1	schizosacch
536	25	61.0	595	1	HMD2_SOLTU	Q41437	solanum tub
537	25	61.0	596	1	HMD1_SOLTU	P48020	solanum tub
538	25	61.0	597	1	NUCD_BUCAP	Q8k9y5	buchnera ap
539	25	61.0	599	1	DX52_HUMAN	Q9y2r4	homo sapien
540	25	61.0	602	1	HMD2_LYCES	P48022	lycopersico
541	25	61.0	604	1	HMD2_CAPAN	Q9xel8	capsicum an
542	25	61.0	604	1	HMDH_NICSY	Q01559	nicotiana s
543	25	61.0	608	1	ALBU_FELCA	P49064	felis silve
544	25	61.0	609	1	GLMS_LEPIN	Q8ezq1	l glucosami
545	25	61.0	610	1	DRTS_TOXGO	Q07422	toxoplasma
546	25	61.0	610	1	RCK2_YEAST	P38623	saccharomyc
547	25	61.0	625	1	BGAL_LACSK	Q48846	lactobacill
548	25	61.0	631	1	YCIQ_ECOLI	P45848	escherichia
549	25	61.0	633	1	Y147_HAEIN	P44543	haemophilus
550	25	61.0	634	1	YKCA_CAEEL	P42083	caenorhabdi
551	25	61.0	639	1	GLGB_BUTFI	P30539	butyrivibri
552	25	61.0	642	1	COG6_CAEEL	Q21270	caenorhabdi
553	25	61.0	646	1	NTP1_HAEPV	O37319	heliothis a
554	25	61.0	647	1	CABC_HUMAN	Q8ni60	homo sapien
555	25	61.0	648	1	NTP1_AMEPV	P29814	amsacta moo
556	25	61.0	649	1	INVA_PHAU	P29001	phaseolus a
557	25	61.0	667	1	UVRB_LACPL	Q88yi8	lactobacill
558	25	61.0	670	1	INV1_MAIZE	P49175	zea mays (m
559	25	61.0	671	1	Z282_HUMAN	Q9udv7	homo sapien
560	25	61.0	679	1	DNLJ_HAEIN	P43813	haemophilus
561	25	61.0	694	1	LCF4_YEAST	P47912	saccharomyc
562	25	61.0	699	1	TOP1_FERIS	O34204	fervidobact
563	25	61.0	703	1	LAGD_LACLA	P59852	lactococcus
564	25	61.0	705	1	GLGB_DEIRA	Q9rtb7	deinococcus
565	25	61.0	722	1	CNA3_HUMAN	Q96by7	homo sapien
566	25	61.0	732	1	YM11_MARPO	P38456	marchantia
567	25	61.0	735	1	VTER_HSV11	P04295	herpes simp
568	25	61.0	736	1	EF2_AERPE	Q9yc19	aeropyrum p
569	25	61.0	747	1	SSP1_BOMMO	P09179	bombyx mori
570	25	61.0	752	1	UGS3_PEA	Q43093	pisum sativ
571	25	61.0	757	1	ECR_LUCCU	O18531	lucilia cup
572	25	61.0	759	1	ARY2_CALVI	P28514	calliphora
573	25	61.0	759	1	FXM1_RAT	P97691	rattus norv
574	25	61.0	763	1	CLPL_LACLA	Q06716	lactococcus
575	25	61.0	787	1	OXAA_CHLMU	Q9pke3	chlamydia m
576	25	61.0	788	1	UGS3_SOLTU	Q43847	solanum tub
577	25	61.0	792	1	SYFB_CHLPN	Q9z7w0	chlamydia p
578	25	61.0	793	1	SYFB_CHLCV	Q824j8	chlamydophi
579	25	61.0	809	1	LEF_BACAN	P15917	bacillus an
580	25	61.0	853	1	MUTS_ECO57	Q9s6p8	escherichia
581	25	61.0	853	1	MUTS_ECOL6	Q8fel3	escherichia
582	25	61.0	853	1	MUTS_ECOLI	P23909	escherichia
583	25	61.0	855	1	MUTS_SALTY	P10339	salmonella
584	25	61.0	904	1	DEGZ_CAEEL	Q10025	caenorhabdi
585	25	61.0	908	1	CNA3_MOUSE	Q80xk6	mus musculo
586	25	61.0	919	1	HEX_ADE12	P19900	human adeno
587	25	61.0	950	1	XRN2_HUMAN	Q9h0d6	homo sapien

588	25	61.0	951	1	XRN2_MOUSE	Q9dbr1	mus musculu
589	25	61.0	952	1	LYAG_HUMAN	P10253	homo sapien
590	25	61.0	953	1	LYAG_MOUSE	P70699	mus musculu
591	25	61.0	969	1	DPOM_NEUIN	P33538	neurospora
592	25	61.0	1003	1	POL5_DROME	Q8i7p9	drosophila
593	25	61.0	1004	1	IMB5_YEAST	P53067	saccharomyc
594	25	61.0	1004	1	YD83_SCHPO	Q10408	schizosacch
595	25	61.0	1012	1	DPOL_HSV6U	P28857	human herpe
596	25	61.0	1020	1	YRD3_CAEEL	Q09573	caenorhabdi
597	25	61.0	1055	1	EPB2_HUMAN	P29323	homo sapien
598	25	61.0	1088	1	PIGO_HUMAN	Q8teq8	homo sapien
599	25	61.0	1088	1	RRPO_ROTBR	P17468	bovine rota
600	25	61.0	1088	1	RRPO_ROTBU	P21615	bovine rota
601	25	61.0	1088	1	RRPO_ROTPTG	P17699	porcine rot
602	25	61.0	1088	1	RRPO_ROTSL	P22678	simian ll r
603	25	61.0	1090	1	SEC5_ARATH	Q8s3u9	arabidopsis
604	25	61.0	1102	1	TR13_YEAST	Q03660	saccharomyc
605	25	61.0	1217	1	S3B3_HUMAN	Q15393	homo sapien
606	25	61.0	1328	1	HUS2_SCHPO	Q09811	schizosacch
607	25	61.0	1342	1	RPOB_BUCAI	P57146	buchnera ap
608	25	61.0	1342	1	RPOB_BUCAP	P41184	buchnera ap
609	25	61.0	1342	1	RPOB_ECOLI	P00575	escherichia
610	25	61.0	1342	1	RPOB_SALTY	P06173	salmonella
611	25	61.0	1355	1	SALM_DROME	P39770	drosophila
612	25	61.0	1376	1	RPOD_ARATH	P56764	arabidopsis
613	25	61.0	1384	1	RPOD_SINAL	Q9thv5	sinapis alb
614	25	61.0	1392	1	RPOB_NEIMA	P57009	neisseria m
615	25	61.0	1392	1	RPOB_NEIMB	Q59622	neisseria m
616	25	61.0	1402	1	SALM_DROVI	P39806	drosophila
617	25	61.0	1405	1	DPOA_SCHPO	P28040	schizosacch
618	25	61.0	1458	1	PHLX_RABIT	Q05017	oryctolagus
619	25	61.0	1490	1	CRK7_HUMAN	Q9nyv4	homo sapien
620	25	61.0	1528	1	SPAA_STRDO	P21979	streptococc
621	25	61.0	1562	1	RPOD_CHLVU	P12465	chlorella v
622	25	61.0	1585	1	YQBO_BACSU	P45931	bacillus su
623	25	61.0	2021	1	OMPA_RICCN	Q52657	rickettsia
624	25	61.0	2145	1	CYAA_PODAN	Q01513	podospora a
625	25	61.0	2249	1	OMPA_RICRI	P15921	rickettsia
626	25	61.0	2278	1	FAB1_YEAST	P34756	saccharomyc
627	25	61.0	2301	1	POLG_TMEVD	P13899	t genome po
628	25	61.0	2303	1	POLG_TMEVB	P08544	t genome po
629	25	61.0	2303	1	POLG_TMEVG	P08545	t genome po
630	25	61.0	2353	1	CCAH_HUMAN	O95180	homo sapien
631	25	61.0	3063	1	CA1C_HUMAN	Q99715	homo sapien
632	25	61.0	3119	1	CA1C_MOUSE	Q60847	mus musculu
633	25	61.0	3898	1	POLG_HCVB	P21530	hog cholera
634	25	61.0	4092	1	DYHC_YEAST	P36022	saccharomyc
635	25	61.0	4349	1	DYHC_FUSSO	P78716	fusarium so
636	25	61.0	4367	1	DYHC_NEUCR	P45443	neurospora
637	25	61.0	4594	1	DYHC_HUMAN	Q14204	homo sapien
638	25	61.0	4644	1	DYHC_MOUSE	Q9jhu4	mus musculu
639	25	61.0	4644	1	DYHC_RAT	P38650	rattus norv
640	25	61.0	4967	1	RYR2_HUMAN	Q92736	homo sapien
641	25	61.0	4969	1	RYR2_RABIT	P30957	oryctolagus
642	24.5	59.8	499	1	UBP6_YEAST	P43593	saccharomyc
643	24.5	59.8	582	1	HEM0_OPSTA	P43090	opsanus tau
644	24	58.5	20	1	CRTC_SPIOL	P30806	spinacia ol



645	24	58.5	68	1	YHBG_KLEOX	P11160	klebsiella
646	24	58.5	76	1	CKS1_PATVU	P41384	patella vul
647	24	58.5	76	1	YB02_ARCFU	O29163	archaeoglob
648	24	58.5	80	1	AFP1_ARATH	P30224	arabidopsis
649	24	58.5	81	1	Y334_ARCFU	O29913	archaeoglob
650	24	58.5	85	1	SIX2_MESMA	Q8i0k7	mesobuthus
651	24	58.5	95	1	MYLE_HUMAN	O95424	homo sapien
652	24	58.5	95	1	MYLE_MOUSE	Q9wuq7	mus musculu
653	24	58.5	106	1	YS73_PSEPK	Q88ix6	pseudomonas
654	24	58.5	112	1	Y145_ARCFU	O30092	archaeoglob
655	24	58.5	120	1	WN1A_ALOVU	P28099	alopias vul
656	24	58.5	120	1	YMG6_RHIME	Q92ug5	rhizobium m
657	24	58.5	124	1	RBFA_BUCAP	Q8k9h2	buchnera ap
658	24	58.5	128	1	Y116_BUCBP	Q89aw2	buchnera ap
659	24	58.5	130	1	YW65_RALSO	Q8xuc6	ralstonia s
660	24	58.5	135	1	NUSB_CLOPE	Q8xjd6	clostridium
661	24	58.5	138	1	YA81_METJA	Q58481	methanococc
662	24	58.5	140	1	YJ14_YEAST	P47111	saccharomyc
663	24	58.5	141	1	HBAB_RANCA	P51465	rana catesb
664	24	58.5	143	1	Y771_METTH	O26865	methanobact
665	24	58.5	148	1	LIMA_RHOER	Q9zag3	rhodococcus
666	24	58.5	151	1	YOZJ_BACSU	O31839	bacillus su
667	24	58.5	152	1	SODC_ZANAE	O65174	zantedeschi
668	24	58.5	157	1	YQAA_HAEIN	P44005	haemophilus
669	24	58.5	162	1	YXII_BACSU	P42301	bacillus su
670	24	58.5	163	1	YC20_SULTO	Q972a7	sulfolobus
671	24	58.5	167	1	YB61_METJA	Q58561	methanococc
672	24	58.5	167	1	YP74_MYCTU	Q50647	mycobacteri
673	24	58.5	171	1	IL26_HUMAN	Q9nph9	homo sapien
674	24	58.5	173	1	LMIP_BOVIN	P20274	bos taurus
675	24	58.5	173	1	YFHC_HAEIN	P44931	haemophilus
676	24	58.5	175	1	CRG2_XENLA	Q91724	xenopus lae
677	24	58.5	175	1	CRG3_XENLA	P55940	xenopus lae
678	24	58.5	180	1	YBYK_CAEEL	P91127	caenorhabdi
679	24	58.5	182	1	YFCM_ECOLI	P76938	escherichia
680	24	58.5	190	1	GTY2_ISSOR	P30102	issatchenki
681	24	58.5	191	1	CYSR_CHICK	P32965	gallus gall
682	24	58.5	191	1	YAF8_CAEEL	P52879	caenorhabdi
683	24	58.5	192	1	CYSR_HUMAN	P21291	homo sapien
684	24	58.5	192	1	CYSR_RAT	P47875	rattus norv
685	24	58.5	192	1	UREE_ALCEU	O30338	alcaligenes
686	24	58.5	193	1	IL18_HUMAN	Q14116	homo sapien
687	24	58.5	194	1	YA58_HAEIN	P44106	haemophilus
688	24	58.5	196	1	MAD2_YEAST	P40958	saccharomyc
689	24	58.5	196	1	YB05_THEAC	Q9hj67	thermoplasm
690	24	58.5	201	1	AlAG_RABIT	P25227	oryctolagus
691	24	58.5	203	1	CHPE_PSEAE	O87005	pseudomonas
692	24	58.5	210	1	ACUA_BACSU	P39065	bacillus su
693	24	58.5	210	1	MOB1_SCHPO	O94360	schizosacch
694	24	58.5	213	1	YOUB_CAEEL	P34658	caenorhabdi
695	24	58.5	214	1	RS6E_SULSO	Q980a6	sulfolobus
696	24	58.5	215	1	MSRA_PSESM	Q88ai5	pseudomonas
697	24	58.5	217	1	MSRA_CORML	Q9apy4	corynebacte
698	24	58.5	218	1	MSRA_BRUME	Q8yde7	brucella me
699	24	58.5	218	1	MSRA_BRUSU	Q8fuz0	brucella su
700	24	58.5	219	1	DEDA_ECOLI	P09548	escherichia
701	24	58.5	219	1	DKA1_YEAST	P14306	saccharomyc

702	24	58.5	222	1	ERG2_YEAST	P32352	saccharomyc
703	24	58.5	222	1	MSA1_SYNY3	P72622	synechocyst
704	24	58.5	222	1	RNS3_ARATH	P42815	arabidopsis
705	24	58.5	224	1	VLT3_VACCV	P07609	vaccinia vi
706	24	58.5	225	1	VLT3_FOWPV	Q9j566	fowlpox vir
707	24	58.5	229	1	YPDP_BACSU	P54163	bacillus su
708	24	58.5	230	1	PURQ_METJA	Q59042	methanococc
709	24	58.5	233	1	MSRA_BOVIN	P54149	bos taurus
710	24	58.5	233	1	MTRP_HUMAN	Q15012	homo sapien
711	24	58.5	233	1	YEG7_YEAST	P39983	saccharomyc
712	24	58.5	235	1	MSRA_HUMAN	Q9uj68	homo sapien
713	24	58.5	236	1	MTAP_SULSO	P50389	sulfolobus
714	24	58.5	240	1	YHBG_ECOLI	P31220	escherichia
715	24	58.5	241	1	Y206_METKA	Q8tyt6	methanopyru
716	24	58.5	246	1	ISPD_CLOTE	Q890ml	clostridium
717	24	58.5	246	1	NPD_THEMA	Q9wyw0	thermotoga
718	24	58.5	254	1	YC8A_METJA	P81318	methanococc
719	24	58.5	255	1	TATC_AZOCH	P54085	azotobacter
720	24	58.5	256	1	TATC_HAEIN	P44560	haemophilus
721	24	58.5	257	1	6PGL_SCHPO	O74455	schizosacch
722	24	58.5	258	1	RSFA_BACSU	P39650	bacillus su
723	24	58.5	259	1	MOTA_TREPA	O07886	treponema p
724	24	58.5	261	1	YM91_ARCFU	O27993	archaeoglob
725	24	58.5	264	1	ME24_SCHPO	Q96ws1	schizosacch
726	24	58.5	267	1	MLF1_MOUSE	Q9qvw4	mus musculus
727	24	58.5	269	1	FLIR_BORBU	Q44907	borrelia bu
728	24	58.5	273	1	LACG_AGRRD	P29824	agrobacteri
729	24	58.5	274	1	YAEF_ECOLI	P37056	escherichia
730	24	58.5	276	1	YE36_AQUAE	O67427	aquifex aeo
731	24	58.5	278	1	ERA_BUCAP	Q8k9r2	buchnera ap
732	24	58.5	279	1	COD5_YEAST	P53195	saccharomyc
733	24	58.5	280	1	YE78_STRP3	Q8k657	streptococc
734	24	58.5	280	1	YG98_STRPY	Q99yh7	streptococc
735	24	58.5	280	1	YH09_STRP8	Q8nzw1	streptococc
736	24	58.5	281	1	SPEE_PYRFU	Q8u4g1	pyrococcus
737	24	58.5	282	1	DRN1_CHICK	Q9ygi5	gallus gall
738	24	58.5	284	1	LEP_PSEFL	P26844	pseudomonas
739	24	58.5	286	1	DAFF_OCEIH	Q8enx2	oceanobacil
740	24	58.5	289	1	TYSY_HSVE2	Q89940	equine herp
741	24	58.5	292	1	GALU_MYCGE	P47691	mycoplasma
742	24	58.5	295	1	YF67_ARCFU	O28705	archaeoglob
743	24	58.5	298	1	ADT3_BOVIN	P32007	bos taurus
744	24	58.5	298	1	ADT3_HUMAN	P12236	homo sapien
745	24	58.5	298	1	Y347_HELPY	O25114	helicobacte
746	24	58.5	299	1	MIME_BOVIN	P19879	bos taurus
747	24	58.5	299	1	YJG8_YEAST	P40363	saccharomyc
748	24	58.5	306	1	CBPA_ECOLI	P36659	escherichia
749	24	58.5	308	1	NOD1_RHIME	P03031	rhizobium m
750	24	58.5	309	1	ZIP2_HUMAN	Q9np94	homo sapien
751	24	58.5	311	1	CYF_SYNEL	Q9x9t1	synechococc
752	24	58.5	312	1	MAB1_STRPN	P35593	s peptide m
753	24	58.5	312	1	TYSY_HUMAN	P04818	homo sapien
754	24	58.5	314	1	MDH_RICCN	Q92ia0	rickettsia
755	24	58.5	316	1	O2K2_HUMAN	Q8ngt1	homo sapien
756	24	58.5	319	1	CDK4_XENLA	Q91727	xenopus lae
757	24	58.5	320	1	MDH_RHIME	Q9eyj6	rhizobium m
758	24	58.5	322	1	NUOH_BUCBP	Q89au0	buchnera ap

759	24	58.5	325	1	NUOH_ECOLI	P33603	escherichia
760	24	58.5	325	1	NUOH_SALTY	Q60010	salmonella
761	24	58.5	326	1	MRAY_STRPN	Q9zha5	streptococc
762	24	58.5	326	1	MRAY_STRR6	Q8dr69	streptococc
763	24	58.5	327	1	THIL_METTH	O27447	methanobact
764	24	58.5	328	1	CEBB_CHICK	Q05826	gallus gall
765	24	58.5	328	1	HEM2_BACHD	Q9k8g2	bacillus ha
766	24	58.5	328	1	YG19_YEAST	P53208	saccharomyc
767	24	58.5	332	1	C1B3_CAVPO	Q9qzz0	cavia porce
768	24	58.5	332	1	NHR9_CAEEL	Q23489	caenorhabdi
769	24	58.5	333	1	Y591_AQUAE	O66853	aquifex aeo
770	24	58.5	336	1	USG_AZOVI	P96199	azotobacter
771	24	58.5	338	1	PHND_ECOLI	P16682	escherichia
772	24	58.5	338	1	RSMC_BUCAI	P57413	buchnera ap
773	24	58.5	340	1	ARGC_STRMU	P59311	streptococc
774	24	58.5	340	1	ILVC_LACLA	Q02138	lactococcus
775	24	58.5	340	1	RIR2_HSV11	P10224	herpes simp
776	24	58.5	340	1	RIR2_HSV1K	P06474	herpes simp
777	24	58.5	341	1	ARGC_STAEP	Q8cp35	staphylococ
778	24	58.5	341	1	LPXK_RHILO	P58185	rhizobium l
779	24	58.5	343	1	Y4FJ_RHISN	P55448	rhizobium s
780	24	58.5	344	1	YA22_AQUAE	O67134	aquifex aeo
781	24	58.5	346	1	FMLR_GORGO	P79176	gorilla gor
782	24	58.5	346	1	FMLR_MACMU	P79189	macaca mula
783	24	58.5	346	1	FMLR_PANTR	P79241	pan troglod
784	24	58.5	346	1	FMLR_PONPY	P79235	pongo pygma
785	24	58.5	346	1	LTAE_PSEAE	Q9htf1	pseudomonas
786	24	58.5	346	1	LTAE_PSESP	O50584	pseudomonas
787	24	58.5	347	1	UL33_HSV6U	P52380	human herpe
788	24	58.5	350	1	FMLR_HUMAN	P21462	homo sapien
789	24	58.5	351	1	FML1_MOUSE	O08790	mus musculu
790	24	58.5	352	1	FMLR_RABIT	Q05394	oryctolagus
791	24	58.5	352	1	OPSB_ORYLA	P87365	oryzias lat
792	24	58.5	360	1	MRAY_HAEDU	Q7vp57	haemophilus
793	24	58.5	364	1	FMLR_MOUSE	P33766	mus musculu
794	24	58.5	365	1	LMC1_MOUSE	Q8veel	mus musculu
795	24	58.5	366	1	MRAY_COREF	Q8fnt7	corynebacte
796	24	58.5	366	1	MRAY_CORGL	Q8nnn2	corynebacte
797	24	58.5	367	1	RCL1_YEAST	Q08096	saccharomyc
798	24	58.5	367	1	SLBP_CAEEL	Q09599	caenorhabdi
799	24	58.5	373	1	CML1_HUMAN	Q99788	homo sapien
800	24	58.5	374	1	MTB1_BACBR	P34905	bacillus br
801	24	58.5	374	1	TGT_YERPE	Q8zc33	yersinia pe
802	24	58.5	374	1	Y006_BORBU	O51039	borrelia bu
803	24	58.5	375	1	TGT_THETN	Q8ram9	thermoanaer
804	24	58.5	377	1	HYPD_RHOCA	P26411	rhodobacter
805	24	58.5	379	1	MTS2_SHISO	P34879	shigella so
806	24	58.5	382	1	GVC1_HALN1	P24574	halobacteri
807	24	58.5	382	1	TGT_LACLA	Q9cj54	lactococcus
808	24	58.5	385	1	TGT_ZYMMO	P28720	zymomonas m
809	24	58.5	386	1	YGIC_ECOLI	P24196	escherichia
810	24	58.5	387	1	PSD6_ARATH	Q93y35	arabidopsis
811	24	58.5	387	1	RH18_SCHPO	O74747	schizosacch
812	24	58.5	389	1	SIR2_MOUSE	Q8vdq8	mus musculu
813	24	58.5	390	1	EMRA_HAEIN	P44928	haemophilus
814	24	58.5	394	1	RLB_ALIPL	P34767	alisma plan
815	24	58.5	395	1	PGK_BRUSU	Q8fyx8	brucella su

816	24	58.5	396	1	PGK_BRUAB	Q91560	brucella ab
817	24	58.5	396	1	PGK_BRUME	Q8yiy0	brucella me
818	24	58.5	397	1	FTSY_BUCBP	Q89b28	buchnera ap
819	24	58.5	398	1	DXR_BUCAP	Q8k9s7	buchnera ap
820	24	58.5	398	1	PNCB_XANCP	Q8pcp3	xanthomonas
821	24	58.5	399	1	ASSY_METTH	O27322	methanobact
822	24	58.5	399	1	HM39_CAEEL	Q22812	caenorhabdi
823	24	58.5	400	1	ENPL_MESAU	P08712	mesocricetu
824	24	58.5	401	1	PNCB_YERPE	Q8zg93	yersinia pe
825	24	58.5	402	1	CAR1_CANPA	P32951	candida par
826	24	58.5	404	1	Y4XM_RHISN	P55705	rhizobium s
827	24	58.5	405	1	HS47_CHICK	P13731	gallus gall
828	24	58.5	407	1	PEPT_STRPN	Q97r31	streptococc
829	24	58.5	412	1	ODP2_RICCN	Q92hk7	rickettsia
830	24	58.5	412	1	Y360_MYCPN	P75241	mycoplasma
831	24	58.5	413	1	MAAL_CLOTT	Q05514	clostridium
832	24	58.5	415	1	ANM3_MOUSE	Q922h1	mus musculu
833	24	58.5	415	1	CRTC_RICCO	P93508	ricinus com
834	24	58.5	415	1	GLYA_LACLA	Q9chw7	lactococcus
835	24	58.5	416	1	CRTC_BERST	Q9zpp1	berberis st
836	24	58.5	416	1	CRTC_BETVU	O81919	beta vulgar
837	24	58.5	416	1	YEBS_HAEIN	P44287	haemophilus
838	24	58.5	417	1	PE23_BOVIN	P34979	bos taurus
839	24	58.5	418	1	CRTC_RABIT	P15253	oryctolagus
840	24	58.5	421	1	CRTC_PRUAR	Q9xf98	prunus arme
841	24	58.5	421	1	RBL_AEGCR	P25413	aegilops cr
842	24	58.5	421	1	RBL_AEGTA	P25414	aegilops ta
843	24	58.5	421	1	TRAB_AGR5	Q44351	agrobacteri
844	24	58.5	422	1	CG2A_SPISO	P04962	spisula sol
845	24	58.5	422	1	RA52_CHICK	P39022	gallus gall
846	24	58.5	424	1	HDA3_MOUSE	O88895	mus musculu
847	24	58.5	425	1	HISX_HALN1	Q9hbw5	halobacteri
848	24	58.5	426	1	RBL_HORVU	P05698	hordeum vul
849	24	58.5	428	1	HDA3_CHICK	P56520	gallus gall
850	24	58.5	428	1	HDA3_HUMAN	O15379	homo sapien
851	24	58.5	429	1	PUR2_VIBCH	Q9kv81	vibrio chol
852	24	58.5	430	1	MESJ_HAEIN	P44689	haemophilus
853	24	58.5	430	1	SERC_ARATH	Q96255	arabidopsis
854	24	58.5	431	1	SPAL_SALTY	P39444	salmonella
855	24	58.5	431	1	T232_BACTB	Q99335	bacillus th
856	24	58.5	432	1	GSAB_LISMO	Q8y6j9	listeria mo
857	24	58.5	434	1	PHD1_SCHPO	O13298	schizosacch
858	24	58.5	434	1	PIGU_MOUSE	Q8k358	mus musculu
859	24	58.5	435	1	PIGU_HUMAN	Q9h490	homo sapien
860	24	58.5	436	1	SYD_HALN1	O07683	halobacteri
861	24	58.5	439	1	YZ04_METJA	Q60260	methanococc
862	24	58.5	440	1	RBL_BAMGL	P51994	bambusa gla
863	24	58.5	443	1	RBL_ABIMR	O78261	abies marie
864	24	58.5	443	1	RBL_IRIGE	Q37227	iris german
865	24	58.5	444	1	CYB_RHOSH	Q02761	rhodobacter
866	24	58.5	444	1	RBL_WATAN	P93936	watsonia an
867	24	58.5	447	1	NISK_LACLA	P42707	lactococcus
868	24	58.5	447	1	RBL_CAMLE	Q95694	camassia le
869	24	58.5	448	1	DHE4_HELPJ	Q9zkd8	helicobacte
870	24	58.5	448	1	DHE4_HELPY	P55990	helicobacte
871	24	58.5	449	1	RBL_ASPEL	P92445	aspidistra
872	24	58.5	449	1	RBL_LIRPL	P93913	liriope pla

873	24	58.5	449	1	RBL_ZAMZA	O98681	zamioculcas
874	24	58.5	450	1	VGLM_HSVEB	P28948	equine herp
875	24	58.5	451	1	GSHR_PSEAE	P23189	pseudomonas
876	24	58.5	451	1	RBL_ARIGL	P93890	aristea gla
877	24	58.5	452	1	HEMF_RHOSH	P33770	rhodobacter
878	24	58.5	452	1	HEMN_RHOSH	P95651	rhodobacter
879	24	58.5	452	1	MENE_HAEIN	P44565	haemophilus
880	24	58.5	452	1	RBL_SALPE	P31201	salvadora p
881	24	58.5	453	1	WAPA_STRMU	P11000	streptococc
882	24	58.5	456	1	GLMU_ECOLI	P17114	escherichia
883	24	58.5	456	1	RBL_SCISC	Q33162	scilla scil
884	24	58.5	457	1	HEMN_ECOLI	P32131	escherichia
885	24	58.5	457	1	HEMN_SALTY	P37129	salmonella
886	24	58.5	457	1	MURC_THEMEA	Q9wy73	thermotoga
887	24	58.5	458	1	RBL_GLAGU	P93906	gladiolus g
888	24	58.5	459	1	RBL_NYPFR	P28261	nypa frutic
889	24	58.5	459	1	RBL_STRLC	Q36800	streptopus
890	24	58.5	460	1	HEMO_RABIT	P20058	oryctolagus
891	24	58.5	460	1	YAGG_ECOLI	P75683	escherichia
892	24	58.5	464	1	GAR3_RAT	P50573	rattus norv
893	24	58.5	465	1	RBL_COROB	Q32040	cornus obli
894	24	58.5	465	1	RBL_DILIN	P28400	dillenia in
895	24	58.5	465	1	VATB_METJA	Q57669	methanococc
896	24	58.5	466	1	RBL_ISOTA	P92463	isophysis t
897	24	58.5	466	1	RBL_MOROL	P48708	moringa ole
898	24	58.5	466	1	RBL_OXADI	P28436	oxalis dill
899	24	58.5	466	1	RBL_TROMA	P48717	tropaeolum
900	24	58.5	467	1	MM18_XENLA	O13065	xenopus lae
901	24	58.5	467	1	RBL_CALUS	P25829	calamus usi
902	24	58.5	467	1	RBL_ERICA	P28413	eriodictyon
903	24	58.5	467	1	RBL_HYDVI	P28425	hydrophyllu
904	24	58.5	467	1	RBL_PHORE	P28262	phoenix rec
905	24	58.5	467	1	RBL_SERRE	P25836	serenoa rep
906	24	58.5	469	1	RBL_AKABI	Q07281	akania bidw
907	24	58.5	469	1	RBL_IRIEN	P92306	iris ensata
908	24	58.5	469	1	RECJ_METJA	Q58387	methanococc
909	24	58.5	471	1	RBL_DRYSU	P28259	drymophloeu
910	24	58.5	471	1	SYE_ECO57	Q8xbn2	escherichia
911	24	58.5	471	1	SYE_ECOL6	Q8ffc9	escherichia
912	24	58.5	471	1	SYE_ECOLI	P04805	escherichia
913	24	58.5	471	1	SYE_SALTY	Q8z4x1	salmonella
914	24	58.5	471	1	SYE_SHIFL	Q83k84	shigella fl
915	24	58.5	472	1	SPSB_BACSU	P39622	bacillus su
916	24	58.5	472	1	SYE_PHOLL	Q7n6y2	photorhabdu
917	24	58.5	472	1	UXAC_BACHD	Q9kez4	bacillus ha
918	24	58.5	472	1	UXAC_LACLA	Q9cf53	lactococcus
919	24	58.5	474	1	RBL_MEDSA	P04991	medicago sa
920	24	58.5	474	1	SYE_VIBCH	O31153	vibrio chol
921	24	58.5	474	1	SYE_VIBPA	Q87rl6	vibrio para
922	24	58.5	475	1	RBL_ABIMA	P24671	abies magni
923	24	58.5	475	1	RBL_CARPA	P48688	carica papa
924	24	58.5	475	1	RBL_CYAPA	P24312	cyanophora
925	24	58.5	476	1	RBL_MAIZE	P00874	zea mays (m
926	24	58.5	476	1	RBL_SETIT	P56647	setaria ita
927	24	58.5	477	1	RBL_AVESA	P48684	avena sativ
928	24	58.5	477	1	RBL_HYOLA	Q9ba49	hyophorbe l
929	24	58.5	477	1	RBL_ORYSA	P12089	oryza sativ

930	24	58.5	477	1	RBL_WHEAT	P11383	triticum ae
931	24	58.5	479	1	RBL_ANACO	P48683	ananas como
932	24	58.5	479	1	RBL_ARATH	O03042	arabidopsis
933	24	58.5	479	1	RBL_BRAOL	P48686	brassica ol
934	24	58.5	480	1	SYE_PASMU	P57906	pasteurella
935	24	58.5	481	1	RNF9_PANTR	Q7yr32	pan troglod
936	24	58.5	482	1	RNF9_HUMAN	Q9udy6	homo sapien
937	24	58.5	487	1	RA18_YEAST	P10862	saccharomyc
938	24	58.5	493	1	FUT3_ARATH	Q9ca71	arabidopsis
939	24	58.5	496	1	AGP2_PIG	Q9bdy7	sus scrofa
940	24	58.5	498	1	YAOI_SCHPO	Q10097	schizosacch
941	24	58.5	502	1	C72J_ARATH	Q9ltm4	arabidopsis
942	24	58.5	502	1	C72K_ARATH	Q9ltm3	arabidopsis
943	24	58.5	504	1	C6AL_DROME	Q9v774	drosophila
944	24	58.5	504	1	GUNW_ERWCA	Q59395	erwinia car
945	24	58.5	506	1	MATK_LEDPA	O62992	ledum palus
946	24	58.5	506	1	PCCB_BACSU	P54541	bacillus su
947	24	58.5	507	1	MATK_CANSA	Q95by0	cannabis sa
948	24	58.5	507	1	MATK_HUMLU	Q95bx9	humulus lup
949	24	58.5	507	1	RECN_CAMJE	Q9phm5	campylobact
950	24	58.5	507	1	YG46_YEAST	P53301	saccharomyc
951	24	58.5	509	1	APR1_ORYSA	Q42456	oryza sativ
952	24	58.5	510	1	ATPA_RHORU	P05036	rhodospiril
953	24	58.5	511	1	G6PD_SOLTU	P37830	solanum tub
954	24	58.5	511	1	VAB2_BOVIN	P31408	bos taurus
955	24	58.5	511	1	VAB2_HUMAN	P21281	homo sapien
956	24	58.5	511	1	VAB2_MOUSE	P50517	mus musculu
957	24	58.5	512	1	ANM3_HUMAN	O60678	homo sapien
958	24	58.5	513	1	ATPA_PASMU	Q9ckw2	pasteurella
959	24	58.5	513	1	VAB1_BOVIN	P31407	bos taurus
960	24	58.5	513	1	VAB1_HUMAN	P15313	homo sapien
961	24	58.5	513	1	VATB_NEUCR	P11593	neurospora
962	24	58.5	514	1	QACA_STAAM	P23215	staphylococ
963	24	58.5	515	1	G6PD_MEDSA	Q42919	medicago sa
964	24	58.5	515	1	GPD6_ARATH	Q9fji5	arabidopsis
965	24	58.5	516	1	GPD5_ARATH	Q9lk23	arabidopsis
966	24	58.5	518	1	YL92_YEAST	P18634	saccharomyc
967	24	58.5	519	1	ENGA_CORGL	Q8nqk6	corynebacte
968	24	58.5	520	1	LAC1_TRAHI	Q02497	trametes hi
969	24	58.5	520	1	LAC1_TRAVI	Q99044	trametes vi
970	24	58.5	520	1	MM28_HUMAN	Q9h239	homo sapien
971	24	58.5	525	1	UBIB_RALSO	Q8y275	ralstonia s
972	24	58.5	527	1	IPT1_YEAST	P38954	saccharomyc
973	24	58.5	528	1	ANM3_RAT	O70467	rattus norv
974	24	58.5	529	1	LAC1_PLEOS	Q12729	pleurotus o
975	24	58.5	530	1	HYAL_LYTVA	O96530	lytechinus
976	24	58.5	530	1	NTG2_HUMAN	Q96cw9	homo sapien
977	24	58.5	530	1	VAT_MOUSE	O35304	mus musculu
978	24	58.5	530	1	VAT_RAT	Q62666	rattus norv
979	24	58.5	533	1	LAC2_PLEOS	Q12739	pleurotus o
980	24	58.5	538	1	PYRG_HELPJ	Q9zm99	helicobacte
981	24	58.5	541	1	UL21_VZVD	P09289	varicella-z
982	24	58.5	546	1	CHOD_STRSQ	P12676	streptomyce
983	24	58.5	547	1	YAGW_ECOLI	P77694	escherichia
984	24	58.5	548	1	GIP2_YEAST	P40036	saccharomyc
985	24	58.5	553	1	G6PI_MYCTU	P77895	mycobacteri
986	24	58.5	554	1	NBL4_MOUSE	P52963	mus musculu

987	24	58.5	555	1	DHAB_CITFR	P45514 citrobacter
988	24	58.5	556	1	NU2M_PODAN	P15578 podospira a
989	24	58.5	563	1	YHJW_ECOLI	P37661 escherichia
990	24	58.5	563	1	YHJW_SALTY	P43666 salmonella
991	24	58.5	565	1	YD08_MYCPN	P75472 mycoplasma
992	24	58.5	567	1	Y391_HUMAN	O15091 homo sapien
993	24	58.5	569	1	PUR1_SOYBN	P52418 glycine max
994	24	58.5	573	1	SECD_MYCTU	Q50634 mycobacteri
995	24	58.5	575	1	HMD1_HEVBR	P29057 hevea brasi
996	24	58.5	575	1	SMF1_YEAST	P38925 saccharomyc
997	24	58.5	575	1	TERM_ADEG1	Q64752 avian adeno
998	24	58.5	575	1	TREZ_ARTRM	Q9ajn6 arthrobacte
999	24	58.5	576	1	IL1R_MOUSE	P13504 mus musculu
1000	24	58.5	577	1	G6PC_SOLTU	Q43839 solanum tub

# ALIGNMENTS

## RESULT 1

### A4\_URDMA

ID A4\_URDMA STANDARD; PRT; 57 AA.  
AC Q29149;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Alzheimer's disease amyloid A4 protein homolog [Contains: Beta-amyloid  
DE protein (Beta-APP) (A-beta)] (Fragment).  
GN APP.  
OS Ursus maritimus (Polar bear) (Thalarctos maritimus).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Ursidae; Ursus.  
OX NCBI\_TaxID=29073;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=92017079; PubMed=1656157;  
RA Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;  
RT "Conservation of the sequence of the Alzheimer's disease amyloid  
RT peptide in dog, polar bear and five other mammals by cross-species  
RT polymerase chain reaction analysis.";  
RL Brain Res. Mol. Brain Res. 10:299-305(1991).  
CC -!- FUNCTION: Functional neuronal receptor which couples to  
CC intracellular signaling pathway through the GTP-binding protein  
CC G(O) (By similarity).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- SIMILARITY: Belongs to the APP family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X56128; CAA39593.1; -.

DR PIR; B60045; B60045.  
 DR HSSP; P05067; 1BA4.  
 DR InterPro; IPR008155; A4\_APP.  
 DR InterPro; IPR001255; Beta-APP.  
 DR Pfam; PF03494; Beta-APP; 1.  
 DR PROSITE; PS00319; A4\_EXTRA; PARTIAL.  
 DR PROSITE; PS00320; A4\_INTRA; PARTIAL.  
 KW Glycoprotein; Amyloid; Neurone; Transmembrane.  
 FT NON\_TER 1 1  
 FT CHAIN 6 48 BETA-AMYLOID PROTEIN (POTENTIAL).  
 FT DOMAIN <1 33 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 34 57 POTENTIAL.  
 FT NON\_TER 57 57  
 SQ SEQUENCE 57 AA; 6172 MW; 84209D88EBA82DFA CRC64;

Query Match 85.4%; Score 35; DB 1; Length 57;  
 Best Local Similarity 100.0%; Pred. No. 0.73;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVFFAED 7  
 |||||  
 Db 22 LVFFAED 28

## RESULT 2

### A4\_CANFA

ID A4\_CANFA STANDARD; PRT; 58 AA.  
 AC Q28280;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Alzheimer's disease amyloid A4 protein homolog [Contains: Beta-amyloid  
 DE protein (Beta-APP) (A-beta)] (Fragment).  
 GN APP.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=92017079; PubMed=1656157;  
 RA Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;  
 RT "Conservation of the sequence of the Alzheimer's disease amyloid  
 RT peptide in dog, polar bear and five other mammals by cross-species  
 RT polymerase chain reaction analysis."  
 RL Brain Res. Mol. Brain Res. 10:299-305(1991).  
 CC -!- FUNCTION: Functional neuronal receptor which couples to  
 CC intracellular signaling pathway through the GTP-binding protein  
 CC G(O) (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- SIMILARITY: Belongs to the APP family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way



CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X56125; CAA39590.1; -.  
DR HSSP; P05067; 1BA4.  
DR InterPro; IPR008155; A4\_APP.  
DR InterPro; IPR001255; Beta-APP.  
DR Pfam; PF03494; Beta-APP; 1.  
DR PROSITE; PS00319; A4\_EXTRA; PARTIAL.  
DR PROSITE; PS00320; A4\_INTRA; PARTIAL.  
KW Glycoprotein; Amyloid; Neurone; Transmembrane.  
FT NON\_TER 1 1  
FT CHAIN 7 49 BETA-AMYLOID PROTEIN (POTENTIAL).  
FT DOMAIN <1 34 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 35 58 POTENTIAL.  
FT NON\_TER 58 58  
SQ SEQUENCE 58 AA; 6285 MW; 8469D488A2E12DFA CRC64;

Query Match 85.4%; Score 35; DB 1; Length 58;  
Best Local Similarity 100.0%; Pred. No. 0.74;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVFFAED 7  
|||  
Db 23 LVFFAED 29

RESULT 3

A4\_RABIT

ID A4\_RABIT STANDARD; PRT; 58 AA.  
AC Q28748;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Alzheimer's disease amyloid A4 protein homolog [Contains: Beta-amyloid  
DE protein (Beta-APP) (A-beta)] (Fragment).  
GN APP.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=92017079; PubMed=1656157;  
RA Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;  
RT "Conservation of the sequence of the Alzheimer's disease amyloid  
RT peptide in dog, polar bear and five other mammals by cross-species  
RT polymerase chain reaction analysis."  
RL Brain Res. Mol. Brain Res. 10:299-305(1991).  
CC -!- FUNCTION: Functional neuronal receptor which couples to  
CC intracellular signaling pathway through the GTP-binding protein  
CC G(O) (By similarity).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- SIMILARITY: Belongs to the APP family.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X56129; CAA39594.1; -.  
DR HSSP; P05067; 1BA4.  
DR InterPro; IPR008155; A4\_APP.  
DR InterPro; IPR001255; Beta-APP.  
DR Pfam; PF03494; Beta-APP; 1.  
DR PROSITE; PS00319; A4\_EXTRA; PARTIAL.  
DR PROSITE; PS00320; A4\_INTRA; PARTIAL.  
KW Glycoprotein; Amyloid; Neurone; Transmembrane.  
FT NON\_TER 1 1  
FT CHAIN 6 48 BETA-AMYLOID PROTEIN (POTENTIAL).  
FT DOMAIN <1 33 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 34 57 POTENTIAL.  
FT DOMAIN 58 >58 CYTOPLASMIC (POTENTIAL).  
FT NON\_TER 58 58  
SQ SEQUENCE 58 AA; 6300 MW; F434209D88EBA82D CRC64;

Query Match 85.4%; Score 35; DB 1; Length 58;  
Best Local Similarity 100.0%; Pred. No. 0.74;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
| | | | | | |  
Db 22 LVFFAED 28

#### RESULT 4

##### A4\_SHEEP

ID A4\_SHEEP STANDARD; PRT; 58 AA.  
AC Q28757;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Alzheimer's disease amyloid A4 protein homolog [Contains: Beta-amyloid  
DE protein (Beta-APP) (A-beta)] (Fragment).  
GN APP.  
OS Ovis aries (Sheep).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Bovidae; Caprinae; Ovis.  
OX NCBI\_TaxID=9940;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Heart;  
RX MEDLINE=92017079; PubMed=1656157;  
RA Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;  
RT "Conservation of the sequence of the Alzheimer's disease amyloid  
RT peptide in dog, polar bear and five other mammals by cross-species  
RT polymerase chain reaction analysis."  
RL Brain Res. Mol. Brain Res. 10:299-305(1991).

CC    -!- FUNCTION: Functional neuronal receptor which couples to  
CC        intracellular signaling pathway through the GTP-binding protein  
CC        G(O) (By similarity).  
CC    -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC    -!- SIMILARITY: Belongs to the APP family.  
CC    -----  
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC    the European Bioinformatics Institute. There are no restrictions on its  
CC    use by non-profit institutions as long as its content is in no way  
CC    modified and this statement is not removed. Usage by and for commercial  
CC    entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC    or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC    -----  
DR    EMBL; X56130; CAA39595.1; -.  
DR    HSSP; P05067; 1BA4.  
DR    InterPro; IPR008155; A4\_APP.  
DR    InterPro; IPR001255; Beta-APP.  
DR    Pfam; PF03494; Beta-APP; 1.  
DR    PROSITE; PS00319; A4\_EXTRA; PARTIAL.  
DR    PROSITE; PS00320; A4\_INTRA; PARTIAL.  
KW    Glycoprotein; Amyloid; Neurone; Transmembrane.  
FT    NON\_TER           1       1  
FT    CHAIN            6       48       BETA-AMYLOID PROTEIN (POTENTIAL).  
FT    DOMAIN           <1      33       EXTRACELLULAR (POTENTIAL).  
FT    TRANSMEM        34      57       POTENTIAL.  
FT    DOMAIN           58     >58       CYTOPLASMIC (POTENTIAL).  
FT    NON\_TER        58      58  
SQ    SEQUENCE    58 AA; 6300 MW; F434209D88EBA82D CRC64;

Query Match                   85.4%; Score 35; DB 1; Length 58;  
Best Local Similarity   100.0%; Pred. No. 0.74;  
Matches       7; Conservative   0; Mismatches    0; Indels       0; Gaps       0;

QY           1 LVFFAED 7  
              |||||||  
Db           22 LVFFAED 28

# RESULT 5

A4\_BOVIN

ID   A4\_BOVIN           STANDARD;       PRT;       59 AA.  
AC   Q28053;  
DT   01-NOV-1997 (Rel. 35, Created)  
DT   01-NOV-1997 (Rel. 35, Last sequence update)  
DT   30-MAY-2000 (Rel. 39, Last annotation update)  
DE   Alzheimer's disease amyloid A4 protein homolog [Contains: Beta-amyloid  
DE   protein (Beta-APP) (A-beta)] (Fragment).  
GN   APP.  
OS   Bos taurus (Bovine).  
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC   Bovidae; Bovinae; Bos.  
OX   NCBI\_TaxID=9913;  
RN   [1]  
RP   SEQUENCE FROM N.A.  
RC   TISSUE=Brain;

RX MEDLINE=92017079; PubMed=1656157;  
 RA Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;  
 RT "Conservation of the sequence of the Alzheimer's disease amyloid  
 RT peptide in dog, polar bear and five other mammals by cross-species  
 RT polymerase chain reaction analysis.";  
 RL Brain Res. Mol. Brain Res. 10:299-305(1991).  
 CC -!- FUNCTION: Functional neuronal receptor which couples to  
 CC intracellular signaling pathway through the GTP-binding protein  
 CC G(O) (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- SIMILARITY: Belongs to the APP family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X56124; CAA39589.1; -.  
 DR EMBL; X56126; CAA39591.1; -.  
 DR HSSP; P05067; 1BA4.  
 DR InterPro; IPR008155; A4\_APP.  
 DR InterPro; IPR001255; Beta-APP.  
 DR Pfam; PF03494; Beta-APP; 1.  
 DR PROSITE; PS00319; A4\_EXTRA; PARTIAL.  
 DR PROSITE; PS00320; A4\_INTRA; PARTIAL.  
 KW Glycoprotein; Amyloid; Neurone; Transmembrane.  
 FT NON\_TER 1 1  
 FT CHAIN 7 49 BETA-AMYLOID PROTEIN (POTENTIAL).  
 FT DOMAIN <1 34 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 35 58 POTENTIAL.  
 FT DOMAIN 59 >59 CYTOPLASMIC (POTENTIAL).  
 FT NON\_TER 59 59  
 SQ SEQUENCE 59 AA; 6414 MW; F43469D488A2E12D CRC64;

Query Match 85.4%; Score 35; DB 1; Length 59;  
 Best Local Similarity 100.0%; Pred. No. 0.76;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVFFAED 7  
 |||||  
 Db 23 LVFFAED 29

#### RESULT 6

##### A4\_SAISC

ID A4\_SAISC STANDARD; PRT; 751 AA.  
 AC Q95241;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Amyloid beta A4 protein precursor (APP) (Alzheimer's disease amyloid  
 DE protein homolog) [Contains: Soluble APP-alpha (S-APP-alpha); Soluble  
 DE APP-beta (S-APP-beta); C99; Beta-amyloid protein 42 (Beta-APP42);  
 DE Beta-amyloid protein 40 (Beta-APP40); C83; P3(42); P3(40); Gamma-

DE CTF(59) (Gamma-secretase C-terminal fragment 59); Gamma-CTF(57)  
 DE (Gamma-secretase C-terminal fragment 57); Gamma-CTF(50) (Gamma-  
 DE secretase C-terminal fragment 50); C31].  
 GN APP.  
 OS Saimiri sciureus (Common squirrel monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Saimiri.  
 OX NCBI\_TaxID=9521;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney, and Liver;  
 RX MEDLINE=96108492; PubMed=8532114;  
 RA Levy E., Amorim A., Frangione B., Walker L.C.;  
 RT "Beta-amyloid precursor protein gene in squirrel monkeys with  
 RT cerebral amyloid angiopathy."  
 RL Neurobiol. Aging 16:805-808(1995).  
 CC -!- FUNCTION: Functions as a cell surface receptor and performs  
 CC physiological functions on the surface of neurons relevant to  
 CC neurite growth, neuronal adhesion and axonogenesis. Involved in  
 CC cell mobility and transcription regulation through protein-protein  
 CC interactions (By similarity). Can promote transcription activation  
 CC through binding to APBB1/Tip60 and inhibit Notch signaling through  
 CC interaction with Numb (By similarity). Couples to apoptosis-  
 CC inducing pathways such as those mediated by G(O) and JIP (By  
 CC similarity). Inhibits G(O) alpha ATPase activity (By similarity).  
 CC Acts as a kinesin I membrane receptor, mediating the axonal  
 CC transport of beta-secretase and presenilin 1 (By similarity). May  
 CC be involved in copper homeostasis/oxidative stress through copper  
 CC ion reduction. In vitro, copper-metallated APP induces neuronal  
 CC death directly or is potentiated through Cu(II)-mediated low-  
 CC density lipoprotein oxidation (By similarity). Can regulate  
 CC neurite outgrowth through binding to components of the  
 CC extracellular matrix such as heparin and collagen I and IV (By  
 CC similarity). The splice isoforms that contain the BPTI domain  
 CC possess protease inhibitor activity (By similarity).  
 CC -!- FUNCTION: Beta-amyloid peptides are lipophilic metal chelators  
 CC with metal-reducing activity. Bind transient metals such as  
 CC copper, zinc and iron (By similarity).  
 CC -!- FUNCTION: The gamma-CTF peptides as well as the caspase-cleaved  
 CC peptides, including C31, are potent enhancers of neuronal  
 CC apoptosis (By similarity).  
 CC -!- SUBUNIT: Binds, via its C-terminal, to the PID domain of several  
 CC cytoplasmic proteins, including APBB family members, the APBA  
 CC family, MAPK8IP1, and SHC1, Numb and Dab1 (By similarity). Binding  
 CC to Dab1 inhibits its serine phosphorylation (By similarity). Also  
 CC interacts with GPCR-like protein BPP, FPRL1, APPBP1, IB1, KNS2  
 CC (via its TPR domains) (By similarity), APPBP2 (via BaSS) and DDB1.  
 CC In vitro, it binds MAPT via the MT-binding domains (By  
 CC similarity). Associates with microtubules in the presence of ATP  
 CC and in a kinesin-dependent manner (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell surface  
 CC protein that rapidly becomes internalized via clathrin-coated  
 CC pits. During maturation, the immature APP (N-glycosylated in the  
 CC endoplasmic reticulum) moves to the Golgi complex where complete  
 CC maturation occurs (O-glycosylated and sulfated). After alpha-  
 CC secretase cleavage, soluble APP is released into the extracellular  
 CC space and the C-terminal is internalized to endosomes and

lysosomes. Some APP accumulates in secretory transport vesicles leaving the late Golgi compartment and returns to the cell surface. Gamma-CTF(59) peptide is located to both the cytoplasm and nuclei of neurons (By similarity).

CC    -!- ALTERNATIVE PRODUCTS:  
CC       Event=Alternative splicing; Named isoforms=2;  
CC       Comment=Additional isoforms seem to exist;  
CC       Name=APP770;  
CC        IsoId=Q95241-1; Sequence=Displayed;  
CC       Name=APP695;  
CC        IsoId=Q95241-2; Sequence=Not described;

CC    -!- DOMAIN: The basolateral sorting signal (BaSS) is required for sorting of membrane proteins to the basolateral surface of epithelial cells (By similarity).

CC    -!- DOMAIN: The NPXY sequence motif found in many tyrosine-phosphorylated proteins is required for the specific binding of the PID domain. However additional amino acids either N- or C-terminal to the NPXY motif are often required for complete interaction. The PID domain-containing proteins which bind APP require the YENPTY motif for full interaction. These interactions are independent of phosphorylation on the terminal tyrosine residue. The NPXY site is also involved in clathrin-mediated endocytosis (By similarity).

CC    -!- PTM: Proteolytically processed under normal cellular conditions. Cleavage by alpha-secretase or alternatively by beta-secretase leads to generation and extracellular release of soluble APP peptides, S-APP-alpha and S-APP-beta, respectively, and the retention of corresponding membrane-anchored C-terminal fragments, C83 and C99. Subsequent processing of C83 by gamma-secretase yields P3 peptides. This is the major secretory pathway and is nonamyloidogenic. Alternatively, presenilin/nicastrin-mediated gamma-secretase processing of C99 releases the amyloid beta proteins, amyloid-beta 40 (Abeta40) and amyloid-beta 42 (Abeta42), major components of amyloid plaques, and the cytotoxic C-terminal fragments, gamma-CTF(50), gamma-CTF(57) and gamma-CTF(59) (By similarity).

CC    -!- PTM: Proteolytically cleaved by caspases during neuronal apoptosis (By similarity). Cleavage at Asp-720 by either caspase-3, -8 or -9 results in the production of the neurotoxic C31 peptide and the increased production of beta-amyloid peptides (By similarity).

CC    -!- PTM: N- and O-glycosylated (By similarity).

CC    -!- PTM: Phosphorylation in the C-terminal on tyrosine, threonine and serine residues is neuron-specific. Phosphorylation can affect APP processing, neuronal differentiation and interaction with other proteins (By similarity).

CC    -!- MISCELLANEOUS: Chelation of metal ions, notably copper, iron and zinc, can induce histidine-bridging between beta-amyloid molecules resulting in beta-amyloid-metal aggregates (By similarity). Extracellular zinc-binding increases binding of heparin to APP and inhibits collagen-binding (By similarity).

CC    -!- SIMILARITY: Belongs to the APP family.

CC    -!- SIMILARITY: Contains 1 BPTI/Kunitz inhibitor domain.

-----

CC    This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

-----
DR EMBL; S81024; AAD14347.1; -.
DR HSSP; P05067; 1AAP.
DR InterPro; IPR008155; A4_APP.
DR InterPro; IPR008154; A4_extra.
DR InterPro; IPR001255; Beta-APP.
DR InterPro; IPR002223; Kunitz_BPTI.
DR Pfam; PF02177; A4_EXTRA; 1.
DR Pfam; PF03494; Beta-APP; 1.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR PRINTS; PR00203; AMYLOIDA4.
DR PRINTS; PR00759; BASICPTASE.
DR ProDom; PD000222; Kunitz_BPTI; 1.
DR SMART; SM00006; A4_EXTRA; 1.
DR SMART; SM00131; KU; 1.
DR PROSITE; PS00319; A4_EXTRA; 1.
DR PROSITE; PS00320; A4_INTRA; 1.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS50279; BPTI_KUNITZ_2; 1.
KW Apoptosis; Endocytosis; Cell adhesion; Serine protease inhibitor;
KW Coated pits; Neurone; Heparin-binding; Metal-binding; Copper; Iron;
KW Zinc; Signal; Transmembrane; Glycoprotein; Phosphorylation;
KW Proteoglycan; Amyloid; Alternative splicing.
FT SIGNAL 1 17 BY SIMILARITY.
FT CHAIN 18 751 A4 PROTEIN.
FT CHAIN 18 668 SOLUBLE APP-ALPHA (POTENTIAL).
FT CHAIN 18 652 SOLUBLE APP-BETA (POTENTIAL).
FT CHAIN 653 751 C99 (POTENTIAL).
FT CHAIN 653 694 BETA-AMYLOID PROTEIN 42 (POTENTIAL).
FT CHAIN 653 692 BETA-AMYLOID PROTEIN 40 (POTENTIAL).
FT CHAIN 669 751 C83 (POTENTIAL).
FT CHAIN 669 694 P3(42) (POTENTIAL).
FT CHAIN 669 692 P3(40) (POTENTIAL).
FT CHAIN 693 751 GAMMA-CTF(59) (POTENTIAL).
FT CHAIN 695 751 GAMMA-CTF(57) (POTENTIAL).
FT CHAIN 702 751 GAMMA-CTF(50) (POTENTIAL).
FT CHAIN 721 751 C31 (POTENTIAL).
FT DOMAIN 18 680 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 681 704 POTENTIAL.
FT DOMAIN 705 751 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 96 110 HEPARIN-BINDING (BY SIMILARITY).
FT DOMAIN 181 188 ZINC-BINDING (BY SIMILARITY).
FT DOMAIN 291 341 BPTI/KUNITZ INHIBITOR.
FT DOMAIN 316 344 HEPARIN-BINDING (BY SIMILARITY).
FT DOMAIN 363 428 HEPARIN-BINDING (BY SIMILARITY).
FT DOMAIN 504 521 COLLAGEN-BINDING (BY SIMILARITY).
FT DOMAIN 713 732 INTERACTION WITH G(O)-ALPHA
FT (BY SIMILARITY).
FT DOMAIN 230 260 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 274 280 POLY-THR.
FT SITE 144 144 REQUIRED FOR COPPER(II) REDUCTION
FT (BY SIMILARITY).
FT ACT_SITE 301 302 REACTIVE BOND.
FT SITE 652 653 CLEAVAGE (BY BETA-SECRETASE)

```

FT				(BY SIMILARITY).
FT	SITE	653	654	CLEAVAGE (BY CASPASE-6) (BY SIMILARITY).
FT	SITE	668	669	CLEAVAGE (BY ALPHA-SECRETASE)
FT				(BY SIMILARITY).
FT	SITE	685	685	INVOLVED IN FREE RADICAL PROPAGATION
FT				(BY SIMILARITY).
FT	SITE	687	687	INVOLVED IN OXIDATIVE REACTIONS
FT				(BY SIMILARITY).
FT	SITE	692	693	CLEAVAGE (BY GAMMA-SECRETASE; SITE 1)
FT				(BY SIMILARITY).
FT	SITE	694	695	CLEAVAGE (BY GAMMA-SECRETASE; SITE 2)
FT				(BY SIMILARITY).
FT	SITE	701	702	CLEAVAGE (BY GAMMA-SECRETASE; SITE 3)
FT				(BY SIMILARITY).
FT	SITE	705	715	BASOLATERAL SORTING SIGNAL
FT				(BY SIMILARITY).
FT	SITE	720	721	CLEAVAGE (BY CASPASES-3,-6,-8 OR -9)
FT				(BY SIMILARITY).
FT	SITE	738	741	ENDOCYTOSIS SIGNAL.
FT	SITE	740	743	NPXY MOTIF.

Query Match 85.4%; Score 35; DB 1; Length 751;  
 Best Local Similarity 100.0%; Pred. No. 9.6;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
 |||||  
 Db 669 LVFFAED 675

# RESULT 7

A4\_CAVPO  
 ID A4\_CAVPO STANDARD; PRT; 770 AA.  
 AC Q60495; Q60496;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Amyloid beta A4 protein precursor (APP) (ABPP) (Alzheimer's disease  
 DE amyloid protein homolog) [Contains: Soluble APP-alpha (S-APP-alpha);  
 DE Soluble APP-beta (S-APP-beta); CTF-alpha; CTF-beta; Beta-amyloid  
 DE protein 42 (Beta-APP42); Beta-amyloid protein 40 (Beta-APP40); P3(42);  
 DE P3(40); CTF(59) (Gamma-secretase C-terminal fragment 59); Gamma-  
 DE CTF(57) (Gamma-secretase C-terminal fragment 57); C31].  
 GN APP.  
 OS Cavia porcellus (Guinea pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.  
 OX NCBI\_TaxID=10141;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RC TISSUE=Brain, and Liver;  
 RX MEDLINE=97236426; PubMed=9116031;  
 RA Beck M., Mueller D., Bigl V.;  
 RT "Amyloid precursor protein in Guinea pigs - complete cDNA sequence and  
 RT alternative splicing."  
 RL Biochim. Biophys. Acta 1351:17-21(1997).  
 RN [2]



RP INTERACTION OF BETA-APP40 WITH APOE.  
 RX MEDLINE=98007700; PubMed=9349544;  
 RA Martel C.L., Mackic J.B., Matsubara E., Governale S., Miguel C.,  
 RA Miao W., McComb J.G., Frangione B., Ghiso J., Zlokovic B.V.;  
 RT "Isoform-specific effects of apolipoproteins E2, E3, and E4 on  
 RT cerebral capillary sequestration and blood-brain barrier transport of  
 RT circulating Alzheimer's amyloid beta.";  
 RL J. Neurochem. 69:1995-2004(1997).  
 RN [3]  
 RP PROCESSING.  
 RX MEDLINE=20084499; PubMed=10619481;  
 RA Beck M., Brueckner M.K., Holzer M., Kaap S., Pannicke T., Arendt T.,  
 RA Bigl V.;  
 RT "Guinea-pig primary cell cultures provide a model to study expression  
 RT and amyloidogenic processing of endogenous amyloid precursor  
 RT protein.";  
 RL Neuroscience 95:243-254(2000).  
 RN [4]  
 RP GAMMA-SECRETASE PROCESSING.  
 RX MEDLINE=20576391; PubMed=11035007;  
 RA Pinnix I., Musunuru U., Tun H., Sridharan A., Golde T., Eckman C.,  
 RA Ziani-Cherif C., Onstead L., Sambamurti K.;  
 RT "A novel gamma -secretase assay based on detection of the putative  
 RT C-terminal fragment-gamma of amyloid beta protein precursor.";  
 RL J. Biol. Chem. 276:481-487(2001).  
 CC -!- FUNCTION: Functions as a cell surface receptor and performs  
 CC physiological functions on the surface of neurons relevant to  
 CC neurite growth, neuronal adhesion and axonogenesis. Involved in  
 CC cell mobility and transcription regulation through protein-protein  
 CC interactions (By similarity). Can promote transcription activation  
 CC through binding to APBB1/Tip60 and inhibit Notch signaling through  
 CC interaction with Numb (By similarity). Couples to apoptosis-  
 CC inducing pathways such as those mediated by G(O) and JIP (By  
 CC similarity). Inhibits G(O) alpha ATPase activity (By similarity).  
 CC Acts as a kinesin I membrane receptor, mediating the axonal  
 CC transport of beta-secretase and presenilin 1 (By similarity). May  
 CC be involved in copper homeostasis/oxidative stress through copper  
 CC ion reduction (By similarity). In vitro, copper-metallated APP  
 CC induces neuronal death directly or is potentiated through Cu(II)-  
 CC mediated low-density lipoprotein oxidation (By similarity). Can  
 CC regulate neurite outgrowth through binding to components of the  
 CC extracellular matrix such as heparin and collagen I and IV (By  
 CC similarity). The splice isoforms that contain the BPTI domain  
 CC possess protease inhibitor activity (By similarity).  
 CC -!- FUNCTION: Beta-amyloid peptides are lipophilic metal chelators  
 CC with metal-reducing activity. Bind transient metals such as  
 CC copper, zinc and iron. Beta-amyloid peptides bind to lipoproteins  
 CC and apolipoproteins E and J in the CSF and to HDL particles in  
 CC plasma, inhibiting metal-catalyzed oxidation of lipoproteins.  
 CC -!- FUNCTION: Appicans elicit adhesion of neural cells to the  
 CC extracellular matrix and may regulate neurite outgrowth in the  
 CC brain (By similarity).  
 CC -!- FUNCTION: The gamma-CTF peptides as well as the caspase-cleaved  
 CC peptides, including C31, are potent enhancers of neuronal  
 CC apoptosis (By similarity).  
 CC -!- SUBUNIT: Binds, via its C-terminal, to the PID domain of several  
 CC cytoplasmic proteins, including APBB family members, the APBA

family, MAPK8IP1, SHC1 and Numb and Dab1 (By similarity). Also interacts with GPCR-like protein BPP, FPRL1, APPBP1, IB1, KNS2 (via its TPR domains), APPBP2 (via BaSS) and DDB1 (By similarity). Associates with microtubules in the presence of ATP and in a kinesin-dependent manner (By similarity). Soluble Abeta40 binds all three isoforms of APOE, in vitro and in vivo. When lipidated, ApoE3 appears to be the preferred amyloid binding isoform, while the apoE4 isoform-beta-APP40 complex is capable of being transported across the blood-brain barrier.

-!- SUBCELLULAR LOCATION: Type I membrane protein. Cell surface protein that rapidly becomes internalized via clathrin-coated pits (By similarity). During maturation, the immature APP (N-glycosylated in the endoplasmic reticulum) moves to the Golgi complex where complete maturation occurs (O-glycosylated and sulfated) (By similarity). After alpha-secretase cleavage, soluble APP is released into the extracellular space and the C-terminal is internalized to endosomes and lysosomes (By similarity). Some APP accumulates in secretory transport vesicles leaving the late Golgi compartment and returns to the cell surface (By similarity). APP sorts to the basolateral surface in epithelial cells (By similarity).

-!- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=2;

Comment=Additional isoforms, missing exons 7,8 and 15, seem to exist. The L-isoforms, missing exon 15, are referred to as appicans;

Name=APP770;

IsoId=Q60495-1; Sequence=Displayed;

Name=APP695;

IsoId=Q60495-2; Sequence=VSP\_007221, VSP\_007222;

-!- TISSUE SPECIFICITY: Isoform APP695 is the major isoform found in brain. The longer isoforms containing the BPTI domain are predominantly expressed in peripheral organs such as muscle and liver.

-!- INDUCTION: Increased levels during neuronal differentiation.

-!- DOMAIN: The basolateral sorting signal (BaSS) is required for sorting of membrane proteins to the basolateral surface of epithelial cells.

-!- DOMAIN: The NPXY sequence motif found in many tyrosine-phosphorylated proteins is required for the specific binding of the PID domain. However additional amino acids either N- or C-terminal to the NPXY motif are often required for complete interaction. The PID domain-containing proteins which bind APP require the YENPTY motif for full interaction. These interactions are independent of phosphorylation on the terminal tyrosine residue (By similarity). The NPXY site is also involved in clathrin-mediated endocytosis.

-!- PTM: Proteolytically processed under normal cellular conditions. Cleavage by alpha-secretase or alternatively by beta-secretase leads to generation and extracellular release of soluble APP peptides, S-APP-alpha and S-APP-beta, respectively, and the retention of corresponding membrane-anchored C-terminal fragments, CTF-alpha and CTF-beta. Subsequent processing of CTF-alpha by gamma-secretase yields P3 peptides. This is the major secretory pathway and is nonamyloidogenic. Alternatively, presenilin/nicastrin-mediated gamma-secretase processing of CTF-beta releases the amyloid beta proteins, amyloid-beta 40 (Abeta40)

CC and amyloid-beta 42 (Abeta42), major components of amyloid  
 CC plaques, and the corresponding cytotoxic C-terminal fragments  
 CC (CTFs).  
 CC -!- PTM: Proteolytically cleaved by caspase-3 during neuronal  
 CC apoptosis (By similarity).  
 CC -!- PTM: N- and O-glycosylated. O-linkage of chondroitin sulfate to  
 CC the L-APP isoforms produces the APP proteoglycan core proteins,  
 CC the appicans (By similarity).  
 CC -!- PTM: Phosphorylation in the C-terminal on tyrosine, threonine and  
 CC serine residues is neuron-specific (By similarity).  
 CC Phosphorylation can affect APP processing, neuronal  
 CC differentiation and interaction with other proteins.  
 CC -!- PTM: Extracellular binding and reduction of copper, results in a  
 CC corresponding oxidation of Cys-144 and Cys-158, and the formation  
 CC of a disulfide bond (By similarity).  
 CC -!- MISCELLANEOUS: Chelation of metal ions, notably copper, iron and  
 CC zinc, can induce histidine-bridging between beta-amyloid molecules  
 CC resulting in beta-amyloid-metal aggregates.  
 CC -!- SIMILARITY: Belongs to the APP family.  
 CC -!- SIMILARITY: Contains 1 BPTI/Kunitz inhibitor domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X97631; CAA66230.1; -.  
 DR EMBL; X99198; CAA67589.1; -.  
 DR HSSP; P05067; 1BA4.  
 DR InterPro; IPR008155; A4\_APP.  
 DR InterPro; IPR008154; A4\_extra.  
 DR InterPro; IPR002223; Kunitz\_BPTI.  
 DR Pfam; PF00014; Kunitz\_BPTI; 1.  
 DR PRINTS; PR00203; AMYLOIDA4.  
 DR PRINTS; PR00759; BASICPTASE.  
 DR ProDom; PD000222; Kunitz\_BPTI; 1.  
 DR SMART; SM00006; A4\_EXTRA; 1.  
 DR SMART; SM00131; KU; 1.  
 DR PROSITE; PS00319; A4\_EXTRA; 1.  
 DR PROSITE; PS00320; A4\_INTRA; 1.  
 DR PROSITE; PS00280; BPTI\_KUNITZ\_1; 1.  
 DR PROSITE; PS50279; BPTI\_KUNITZ\_2; 1.  
 KW Apoptosis; Endocytosis; Cell adhesion; Serine protease inhibitor;  
 KW Coated pits; Neurone; Heparin-binding; Metal-binding; Copper; Iron;  
 KW Zinc; Signal; Transmembrane; Glycoprotein; Phosphorylation;  
 KW Proteoglycan; Alternative splicing; Amyloid.  
 FT SIGNAL 1 17 BY SIMILARITY.  
 FT CHAIN 18 770 AMYLOID BETA A4 PROTEIN.  
 FT CHAIN 18 687 SOLUBLE APP-ALPHA (BY SIMILARITY).  
 FT CHAIN 18 671 SOLUBLE APP-BETA (BY SIMILARITY).  
 FT CHAIN 672 770 CTF-ALPHA (BY SIMILARITY).  
 FT CHAIN 672 713 BETA-AMYLOID PROTEIN 42 (BY SIMILARITY).  
 FT CHAIN 672 711 BETA-AMYLOID PROTEIN 40 (BY SIMILARITY).  
 FT CHAIN 688 770 CTF-BETA (BY SIMILARITY).

FT	CHAIN	688	713	P3(42) (BY SIMILARITY).
FT	CHAIN	688	711	P3(40) (BY SIMILARITY).
FT	CHAIN	712	770	GAMMA-CTF(59) (BY SIMILARITY).
FT	CHAIN	714	770	GAMMA-CTF(57) (BY SIMILARITY).

Query Match 85.4%; Score 35; DB 1; Length 770;  
 Best Local Similarity 100.0%; Pred. No. 9.8;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVFFAED 7  
 |||||  
 Db 688 LVFFAED 694

# RESULT 8

## A4\_HUMAN

ID A4\_HUMAN STANDARD; PRT; 770 AA.  
 AC P05067; P09000; P78438; Q13764; Q13778; Q13793; Q16011; Q16014;  
 AC Q16019; Q16020; Q9BT38; Q9UCA9; Q9UCB6; Q9UCC8; Q9UCD1; Q9UQ58;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Amyloid beta A4 protein precursor (APP) (ABPP) (Alzheimer's disease  
 DE amyloid protein) (Cerebral vascular amyloid peptide) (CVAP) (Protease  
 DE nexin-II) (PN-II) (APPI) (PreA4) [Contains: Soluble APP-alpha (S-APP-  
 DE alpha); Soluble APP-beta (S-APP-beta); C99; Beta-amyloid protein 42  
 DE (Beta-APP42); Beta-amyloid protein 40 (Beta-APP40); C83; P3(42);  
 DE P3(40); Gamma-CTF(59) (Gamma-secretase C-terminal fragment 59)  
 DE (Amyloid intracellular domain 59) (AID(59)); Gamma-CTF(57) (Gamma-  
 DE secretase C-terminal fragment 57) (Amyloid intracellular domain 57)  
 DE (AID(57)); Gamma-CTF(50) (Gamma-secretase C-terminal fragment 50)  
 DE (Amyloid intracellular domain 50) (AID(50)); C31].  
 GN APP OR A4 OR AD1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM APP695).  
 RC TISSUE=Brain;  
 RX MEDLINE=87144572; PubMed=2881207;  
 RA Kang J., Lemaire H.-G., Unterbeck A., Salbaum J.M., Masters C.L.,  
 RA Grzeschik K.-H., Multhaup G., Beyreuther K., Mueller-Hill B.;  
 RT "The precursor of Alzheimer's disease amyloid A4 protein resembles a  
 RT cell-surface receptor.";  
 RL Nature 325:733-736(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM APP751).  
 RC TISSUE=Brain;  
 RX MEDLINE=88122639; PubMed=2893289;  
 RA Ponte P., Gonzalez-Dewhitt P., Schilling J., Miller J., Hsu D.,  
 RA Greenberg B., Davis K., Wallace W., Lieberburg I., Fuller F.,  
 RA Cordell B.;  
 RT "A new A4 amyloid mRNA contains a domain homologous to serine  
 RT proteinase inhibitors.";  
 RL Nature 331:525-527(1988).  
 RN [3]

RP SEQUENCE FROM N.A. (ISOFORM APP695).  
 RX MEDLINE=89128427; PubMed=2783775;  
 RA Lemaire H.-G., Salbaum J.M., Multhaup G., Kang J., Bayney R.M.,  
 RA Unterbeck A., Beyreuther K., Mueller-Hill B.;  
 RT "The PreA4(695) precursor protein of Alzheimer's disease A4 amyloid  
 RT is encoded by 16 exons.";  
 RL Nucleic Acids Res. 17:517-522(1989).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM APP770).  
 RX MEDLINE=90236318; PubMed=2110105;  
 RA Yoshikai S.-I., Sasaki H., Doh-Ura K., Furuya H., Sakaki Y.;  
 RT "Genomic organization of the human amyloid beta-protein precursor  
 RT gene.";  
 RL Gene 87:257-263(1990).  
 RN [5]  
 RP ERRATUM, AND REVISIONS.  
 RA Yoshikai S.-I., Sasaki H., Doh-ura K., Furuya H., Sakaki Y.;  
 RL Gene 102:291-292(1991).  
 RN [6]  
 RP SEQUENCE FROM N.A. (ISOFORM L-APP733).  
 RC TISSUE=Leukocyte;  
 RX MEDLINE=92268136; PubMed=1587857;  
 RA Koenig G., Moenning U., Czech C., Prior R., Banati R.,  
 RA Schreiter-Gasser U., Bauer J., Masters C.L., Beyreuther K.;  
 RT "Identification and differential expression of a novel alternative  
 RT splice isoform of the beta A4 amyloid precursor protein (APP) mRNA in  
 RT leukocytes and brain microglial cells.";  
 RL J. Biol. Chem. 267:10804-10809(1992).  
 RN [7]  
 RP SEQUENCE FROM N.A. (ISOFORM APP770).  
 RX MEDLINE=97263807; PubMed=9108164;  
 RA Hattori M., Tsukahara F., Furuhashi Y., Tanahashi H., Hirose M.,  
 RA Saito M., Tsukuni S., Sakaki Y.;  
 RT "A novel method for making nested deletions and its application for  
 RT sequencing of a 300 kb region of human APP locus.";  
 RL Nucleic Acids Res. 25:1802-1808(1997).  
 RN [8]  
 RP SEQUENCE FROM N.A. (ISOFORM APP639).  
 RC TISSUE=Brain;  
 RX MEDLINE=22744650; PubMed=12859342;  
 RA Tang K., Wang C., Shen C., Sheng S., Ravid R., Jing N.;  
 RT "Identification of a novel alternative splicing isoform of human  
 RT amyloid precursor protein gene, APP639.";  
 RL Eur. J. Neurosci. 18:102-108(2003).  
 RN [9]  
 RP SEQUENCE FROM N.A. (ISOFORM APP305).  
 RC TISSUE=Pancreas;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnierch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [10]  
 RP SEQUENCE OF 1-10 FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=89016647; PubMed=3140222;  
 RA Schon E.A., Mita S., Sadlock J., Herbert J.;  
 RT "A cDNA specifying the human amyloid beta precursor protein (ABPP)  
 RT encodes a 95-kDa polypeptide.";  
 RL Nucleic Acids Res. 16:9351-9351(1988).  
 RN [11]  
 RP ERRATUM, AND REVISIONS.  
 RA Mita S., Sadlock J., Herbert J., Schon E.A.;  
 RL Nucleic Acids Res. 16:11402-11402(1988).  
 RN [12]  
 RP SEQUENCE OF 1-75 FROM N.A.  
 RX MEDLINE=89165870; PubMed=2538123;  
 RA La Fauci G., Lahiri D.K., Salton S.R., Robakis N.K.;  
 RT "Characterization of the 5'-end region and the first two exons of the  
 RT beta-protein precursor gene.";  
 RL Biochem. Biophys. Res. Commun. 159:297-304(1989).  
 RN [13]  
 RP SEQUENCE OF 18-50.  
 RC TISSUE=Fibroblast;  
 RX MEDLINE=87250462; PubMed=3597385;  
 RA van Nostrand W.E., Cunningham D.D.;  
 RT "Purification of protease nexin II from human fibroblasts.";  
 RL J. Biol. Chem. 262:8508-8514(1987).  
 RN [14]  
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORM APP751).  
 RC TISSUE=Brain;  
 RX MEDLINE=89346754; PubMed=2569763;  
 RA de Sauvage F., Octave J.N.;  
 RT "A novel mRNA of the A4 amyloid precursor gene coding for a possibly  
 RT secreted protein.";  
 RL Science 245:651-653(1989).  
 RN [15]  
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORM APP695).  
 RC TISSUE=Brain;  
 RX MEDLINE=87231971; PubMed=3035574;  
 RA Robakis N.K., Ramakrishna N., Wolfe G., Wisniewski H.M.;  
 RT "Molecular cloning and characterization of a cDNA encoding the  
 RT cerebrovascular and the neuritic plaque amyloid peptides.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:4190-4194(1987).  
 RN [16]  
 RP SEQUENCE OF 286-366 FROM N.A.  
 RX MEDLINE=88122640; PubMed=2893290;  
 RA Tanzi R.E., McClatchey A.I., Lamperti E.D., Villa-Komaroff L.,

RA Gusella J.F., Neve R.L.;  
 RT "Protease inhibitor domain encoded by an amyloid protein precursor  
 RT mRNA associated with Alzheimer's disease.";  
 RL Nature 331:528-530(1988).  
 RN [17]  
 RP SEQUENCE OF 287-367 FROM N.A.  
 RX MEDLINE=88122641; PubMed=2893291;  
 RA Kitaguchi N., Takahashi Y., Tokushima Y., Shiojiri S., Ito H.;  
 RT "Novel precursor of Alzheimer's disease amyloid protein shows  
 RT protease inhibitory activity.";  
 RL Nature 331:530-532(1988).  
 RN [18]  
 RP SEQUENCE OF 507-770 FROM N.A.  
 RC TISSUE=Brain cortex;  
 RX MEDLINE=88124954; PubMed=2893379;  
 RA Zain S.B., Salim M., Chou W.G., Sajdel-Sulkowska E.M., Majocha R.E.,  
 RA Marotta C.A.;  
 RT "Molecular cloning of amyloid cDNA derived from mRNA of the Alzheimer  
 RT disease brain: coding and noncoding regions of the fetal precursor  
 RT mRNA are expressed in the cortex.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 85:929-933(1988).  
 RN [19]  
 RP SEQUENCE OF 523-555, AND COLLAGEN-BINDING DOMAIN.  
 RX MEDLINE=96139497; PubMed=8576160;  
 RA Behr D., Hesse L., Masters C.L., Multhaup G.;  
 RT "Regulation of amyloid protein precursor (APP) binding to collagen and  
 RT mapping of the binding sites on APP and collagen type I.";  
 RL J. Biol. Chem. 271:1613-1620(1996).  
 RN [20]  
 RP SEQUENCE OF 655-737 FROM N.A., AND VARIANTS AD PHE-717; AD ILE-717  
 RP AND AD GLY-717.  
 RX MEDLINE=93236601; PubMed=8476439;  
 RA Denman R.B., Rosenzwaig R., Miller D.L.;  
 RT "A system for studying the effect(s) of familial Alzheimer disease  
 RT mutations on the processing of the beta-amyloid peptide precursor.";  
 RL Biochem. Biophys. Res. Commun. 192:96-103(1993).  
 RN [21]  
 RP SEQUENCE OF 656-737 FROM N.A.  
 RX MEDLINE=89392030; PubMed=2675837;  
 RA Johnstone E.M., Chaney M.O., Moore R.E., Ward K.E., Norris F.H.,  
 RA Little S.P.;  
 RT "Alzheimer's disease amyloid peptide is encoded by two exons and shows  
 RT similarity to soybean trypsin inhibitor.";  
 RL Biochem. Biophys. Res. Commun. 163:1248-1255(1989).  
 RN [22]

Query Match 85.4%; Score 35; DB 1; Length 770;  
 Best Local Similarity 100.0%; Pred. No. 9.8;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
 |||||  
 Db 688 LVFFAED 694

RESULT 9  
 A4\_MACFA

ID A4\_MACFA STANDARD; PRT; 770 AA.  
 AC P53601; Q95KN7;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Amyloid beta A4 protein precursor (APP) (ABPP) (Alzheimer's disease  
 DE amyloid protein homolog) [Contains: Soluble APP-alpha (S-APP-alpha);  
 DE Soluble APP-beta (S-APP-beta); C99; Beta-amyloid protein 42 (Beta-  
 DE APP42); Beta-amyloid protein 40 (Beta-APP40); C83; P3(42); P3(40);  
 DE Gamma-CTF(59) (Gamma-secretase C-terminal fragment 59); Gamma-CTF(57)  
 DE (Gamma-secretase C-terminal fragment 57); Gamma-CTF(50) (Gamma-  
 DE secretase C-terminal fragment 50); C31].  
 GN APP.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;  
 OC Cercopithecinae; Macaca.  
 OX NCBI\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS APP695 AND APP770).  
 RC TISSUE=Cerebellum;  
 RX MEDLINE=91273117; PubMed=1905108;  
 RA Podlisny M.B., Tolan D.R., Selkoe D.J.;  
 RT "Homology of the amyloid beta protein precursor in monkey and human  
 RT supports a primate model for beta amyloidosis in Alzheimer's  
 RT disease.";  
 RL Am. J. Pathol. 138:1423-1435(1991).  
 CC -!- FUNCTION: Functions as a cell surface receptor and performs  
 CC physiological functions on the surface of neurons relevant to  
 CC neurite growth, neuronal adhesion and axonogenesis. Involved in  
 CC cell mobility and transcription regulation through protein-protein  
 CC interactions (By similarity). Can promote transcription activation  
 CC through binding to APBB1/Tip60 and inhibit Notch signaling through  
 CC interaction with Numb (By similarity). Couples to apoptosis-  
 CC inducing pathways such as those mediated by G(O) and JIP (By  
 CC similarity). Inhibits G(O) alpha ATPase activity (By similarity).  
 CC Acts as a kinesin I membrane receptor, mediating the axonal  
 CC transport of beta-secretase and presenilin 1 (By similarity). May  
 CC be involved in copper homeostasis/oxidative stress through copper  
 CC ion reduction. In vitro, copper-metallated APP induces neuronal  
 CC death directly or is potentiated through Cu(II)-mediated low-  
 CC density lipoprotein oxidation (By similarity). Can regulate  
 CC neurite outgrowth through binding to components of the  
 CC extracellular matrix such as heparin and collagen I and IV (By  
 CC similarity). The splice isoforms that contain the BPTI domain  
 CC possess protease inhibitor activity (By similarity).  
 CC -!- FUNCTION: Beta-amyloid peptides are lipophilic metal chelators  
 CC with metal-reducing activity. Bind transient metals such as  
 CC copper, zinc and iron (By similarity).  
 CC -!- FUNCTION: The gamma-CTF peptides as well as the caspase-cleaved  
 CC peptides, including C31, are potent enhancers of neuronal  
 CC apoptosis (By similarity).  
 CC -!- SUBUNIT: Binds, via its C-terminal, to the PID domain of several  
 CC cytoplasmic proteins, including APBB family members, the APBA  
 CC family, MAPK8IP1, and SHC1, Numb and Dab1 (By similarity). Binding  
 CC to Dab1 inhibits its serine phosphorylation (By similarity). Also  
 CC interacts with GPCR-like protein BPP, FPRL1, APPBP1, IB1, KNS2



(via its TPR domains) (By similarity), APPBP2 (via BaSS) and DDB1. In vitro, it binds MAPT via the MT-binding domains (By similarity). Associates with microtubules in the presence of ATP and in a kinesin-dependent manner (By similarity).

CC    -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell surface protein that rapidly becomes internalized via clathrin-coated pits. During maturation, the immature APP (N-glycosylated in the endoplasmic reticulum) moves to the Golgi complex where complete maturation occurs (O-glycosylated and sulfated). After alpha-secretase cleavage, soluble APP is released into the extracellular space and the C-terminal is internalized to endosomes and lysosomes. Some APP accumulates in secretory transport vesicles leaving the late Golgi compartment and returns to the cell surface. Gamma-CTF(59) peptide is located to both the cytoplasm and nuclei of neurons (By similarity).

CC    -!- ALTERNATIVE PRODUCTS:

CC       Event=Alternative splicing; Named isoforms=2;

CC       Comment=Additional isoforms seem to exist;

CC       Name=APP770;

CC       IsoId=P53601-1; Sequence=Displayed;

CC       Name=APP695;

CC       IsoId=P53601-2; Sequence=VSP\_000010, VSP\_000011;

CC    -!- DOMAIN: The basolateral sorting signal (BaSS) is required for sorting of membrane proteins to the basolateral surface of epithelial cells (By similarity).

CC    -!- DOMAIN: The NPXY sequence motif found in many tyrosine-phosphorylated proteins is required for the specific binding of the PID domain. However additional amino acids either N- or C-terminal to the NPXY motif are often required for complete interaction. The PID domain-containing proteins which bind APP require the YENPTY motif for full interaction. These interactions are independent of phosphorylation on the terminal tyrosine residue. The NPXY site is also involved in clathrin-mediated endocytosis (By similarity).

CC    -!- PTM: Proteolytically processed under normal cellular conditions. Cleavage by alpha-secretase or alternatively by beta-secretase leads to generation and extracellular release of soluble APP peptides, S-APP-alpha and S-APP-beta, respectively, and the retention of corresponding membrane-anchored C-terminal fragments, C83 and C99. Subsequent processing of C83 by gamma-secretase yields P3 peptides. This is the major secretory pathway and is nonamyloidogenic. Alternatively, presenilin/nicastrin-mediated gamma-secretase processing of C99 releases the amyloid beta proteins, amyloid-beta 40 (Abeta40) and amyloid-beta 42 (Abeta42), major components of amyloid plaques, and the cytotoxic C-terminal fragments, gamma-CTF(50), gamma-CTF(57) and gamma-CTF(59) (By similarity).

CC    -!- PTM: Proteolytically cleaved by caspases during neuronal apoptosis (By similarity). Cleavage at Asp-739 by either caspase-3, -8 or -9 results in the production of the neurotoxic C31 peptide and the increased production of beta-amyloid peptides (By similarity).

CC    -!- PTM: N- and O-glycosylated (By similarity).

CC    -!- PTM: Phosphorylation in the C-terminal on tyrosine, threonine and serine residues is neuron-specific. Phosphorylation can affect APP processing, neuronal differentiation and interaction with other proteins (By similarity).

CC    -!- MISCELLANEOUS: Chelation of metal ions, notably copper, iron and

CC zinc, can induce histidine-bridging between beta-amyloid molecules  
 CC resulting in beta-amyloid-metal aggregates (By similarity).  
 CC Extracellular zinc-binding increases binding of heparin to APP and  
 CC inhibits collagen-binding (By similarity).  
 CC -!- SIMILARITY: Belongs to the APP family.  
 CC -!- SIMILARITY: Contains 1 BPTI/Kunitz inhibitor domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M58727; AAA36829.1; -.  
 DR EMBL; M58726; AAA36828.1; -.  
 DR HSSP; P05067; 1AAP.  
 DR InterPro; IPR008155; A4\_APP.  
 DR InterPro; IPR008154; A4\_extra.  
 DR InterPro; IPR001255; Beta-APP.  
 DR InterPro; IPR002223; Kunitz\_BPTI.  
 DR Pfam; PF02177; A4\_EXTRA; 1.  
 DR Pfam; PF03494; Beta-APP; 1.  
 DR Pfam; PF00014; Kunitz\_BPTI; 1.  
 DR PRINTS; PR00203; AMYLOIDA4.  
 DR PRINTS; PR00759; BASICPTASE.  
 DR ProDom; PD000222; Kunitz\_BPTI; 1.  
 DR SMART; SM00006; A4\_EXTRA; 1.  
 DR SMART; SM00131; KU; 1.  
 DR PROSITE; PS00319; A4\_EXTRA; 1.  
 DR PROSITE; PS00320; A4\_INTRA; 1.  
 DR PROSITE; PS00280; BPTI\_KUNITZ\_1; 1.  
 DR PROSITE; PS50279; BPTI\_KUNITZ\_2; 1.  
 KW Apoptosis; Endocytosis; Cell adhesion; Serine protease inhibitor;  
 KW Coated pits; Neurone; Heparin-binding; Metal-binding; Copper; Iron;  
 KW Zinc; Signal; Transmembrane; Glycoprotein; Phosphorylation;  
 KW Proteoglycan; Alternative splicing; Amyloid.  
 FT SIGNAL 1 17 BY SIMILARITY.  
 FT CHAIN 18 770 AMYLOID BETA A4 PROTEIN.  
 FT CHAIN 18 687 SOLUBLE APP-ALPHA (POTENTIAL).  
 FT CHAIN 18 671 SOLUBLE APP-BETA (POTENTIAL).  
 FT CHAIN 672 770 C99 (POTENTIAL).  
 FT CHAIN 672 713 BETA-AMYLOID PROTEIN 42 (POTENTIAL).  
 FT CHAIN 672 711 BETA-AMYLOID PROTEIN 40 (POTENTIAL).  
 FT CHAIN 688 770 C83 (POTENTIAL).  
 FT CHAIN 688 713 P3(42) (POTENTIAL).  
 FT CHAIN 688 711 P3(40) (POTENTIAL).  
 FT CHAIN 712 770 GAMMA-CTF(59) (POTENTIAL).  
 FT CHAIN 714 770 GAMMA-CTF(57) (POTENTIAL).  
 FT CHAIN 721 770 GAMMA-CTF(50) (POTENTIAL).  
 FT CHAIN 740 770 C31 (POTENTIAL).  
 FT DOMAIN 18 699 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 700 723 POTENTIAL.  
 FT DOMAIN 724 770 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 96 110 HEPARIN-BINDING (BY SIMILARITY).  
 FT DOMAIN 181 188 ZINC-BINDING (BY SIMILARITY).

FT	DOMAIN	291	341	BPTI/KUNITZ INHIBITOR.
FT	DOMAIN	391	423	HEPARIN-BINDING (BY SIMILARITY).
FT	DOMAIN	491	522	HEPARIN-BINDING (BY SIMILARITY).
FT	DOMAIN	523	540	COLLAGEN-BINDING (BY SIMILARITY).
FT	DOMAIN	732	751	INTERACTION WITH G(O)-ALPHA
FT				(BY SIMILARITY).
FT	DOMAIN	230	260	ASP/GLU-RICH (ACIDIC).
FT	DOMAIN	274	280	POLY-THR.
FT	SITE	144	144	REQUIRED FOR COPPER(II) REDUCTION
FT				(BY SIMILARITY).
FT	ACT_SITE	301	302	REACTIVE BOND (BY SIMILARITY).
FT	SITE	671	672	CLEAVAGE (BY BETA-SECRETASE)
FT				(BY SIMILARITY).
FT	SITE	672	673	CLEAVAGE (BY CASPASE-6) (BY SIMILARITY).
FT	SITE	687	688	CLEAVAGE (BY ALPHA-SECRETASE)
FT				(BY SIMILARITY).
FT	SITE	704	704	IMPLICATED IN FREE RADICAL PROPAGATION
FT				(BY SIMILARITY).
FT	SITE	706	706	INVOLVED IN OXIDATIVE REACTIONS
FT				(BY SIMILARITY).
FT	SITE	711	712	CLEAVAGE (BY GAMMA-SECRETASE; SITE 1)
FT				(BY SIMILARITY).
FT	SITE	713	714	CLEAVAGE (BY GAMMA-SECRETASE; SITE 2)
FT				(BY SIMILARITY).
FT	SITE	720	721	CLEAVAGE (BY GAMMA-SECRETASE; SITE 3)
FT				(BY SIMILARITY).
FT	SITE	724	734	BASOLATERAL SORTING SIGNAL
FT				(BY SIMILARITY).
FT	SITE	739	740	CLEAVAGE (BY CASPASES-3,-6,-8 OR -9)

Query Match 85.4%; Score 35; DB 1; Length 770;  
 Best Local Similarity 100.0%; Pred. No. 9.8;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
 |||||  
 Db 688 LVFFAED 694

# RESULT 10

## A4\_MOUSE

ID A4\_MOUSE STANDARD; PRT; 770 AA.  
 AC P12023; P97487; P97942; Q99K32;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Amyloid beta A4 protein precursor (APP) (ABPP) (Alzheimer's disease  
 DE amyloid protein homolog) (Amyloidogenic glycoprotein) (AG) [Contains:  
 DE Soluble APP-alpha (S-APP-alpha); Soluble APP-beta (S-APP-beta); C99  
 DE (APP-C99); Beta-amyloid protein 42 (Beta-APP42); Beta-amyloid protein  
 DE 40 (Beta-APP40); C83; P3(42); P3(40); Gamma-CTF(59) (Gamma-secretase  
 DE C-terminal fragment 59) (Amyloid intracellular domain 59) (AID(59))  
 DE (APP-C59); Gamma-CTF(57) (Gamma-secretase C-terminal fragment 57)  
 DE (Amyloid intracellular domain 57) (AID(57)) (APP-C57); Gamma-CTF(50)  
 DE (Gamma-secretase C-terminal fragment 50) (Amyloid intracellular domain  
 DE 50) (AID(50)); C31].  
 GN APP.

OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM APP695).  
 RC TISSUE=Brain;  
 RX MEDLINE=88106489; PubMed=3322280;  
 RA Yamada T., Sasaki H., Furuya H., Miyata T., Goto I., Sakaki Y.;  
 RT "Complementary DNA for the mouse homolog of the human amyloid beta  
 RT protein precursor."  
 RL Biochem. Biophys. Res. Commun. 149:665-671(1987).  
 RN [2]  
 RP REVISIONS.  
 RA Yamada T.;  
 RL Submitted (MAR-1988) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM APP695).  
 RC STRAIN=BALB/c; TISSUE=Brain;  
 RX MEDLINE=92096458; PubMed=1756177;  
 RA de Strooper B., van Leuven F., van den Berghe H.;  
 RT "The amyloid beta protein precursor or proteinase nexin II from mouse  
 RT is closer related to its human homolog than previously reported."  
 RL Biochim. Biophys. Acta 1129:141-143(1991).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM APP695).  
 RC STRAIN=SAMP8; TISSUE=Hippocampus;  
 RX MEDLINE=21130647; PubMed=11235921;  
 RA Kumar V.B., Vyas K., Franko M., Choudhary V., Buddhiraju C.,  
 RA Alvarez J., Morley J.E.;  
 RT "Molecular cloning, expression, and regulation of hippocampal amyloid  
 RT precursor protein of senescence accelerated mouse (SAMP8).";  
 RL Biochem. Cell Biol. 79:57-67(2001).  
 RN [5]  
 RP SEQUENCE OF 1-19 FROM N.A.  
 RX MEDLINE=92209998; PubMed=1555768;  
 RA Izumi R., Yamada T., Yoshikai S.I., Sasaki H., Hattori M.,  
 RA Sakai Y.;  
 RT "Positive and negative regulatory elements for the expression of the  
 RT Alzheimer's disease amyloid precursor-encoding gene in mouse."  
 RL Gene 112:189-195(1992).  
 RN [6]  
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORM APP770).  
 RC TISSUE=Breast tumor;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [7]  
 RP SEQUENCE OF 281-380 FROM N.A., AND ALTERNATIVE SPLICING.  
 RC TISSUE=Brain, and Kidney;  
 RX MEDLINE=89149813; PubMed=2493250;  
 RA Yamada T., Sasaki H., Dohura K., Goto I., Sakaki Y.;  
 RT "Structure and expression of the alternatively-spliced forms of mRNA  
 RT for the mouse homolog of Alzheimer's disease amyloid beta protein  
 RT precursor.";  
 RL Biochem. Biophys. Res. Commun. 158:906-912(1989).  
 RN [8]  
 RP SEQUENCE OF 289-364 FROM N.A.  
 RC STRAIN=CD-1; TISSUE=Placenta;  
 RX MEDLINE=89345111; PubMed=2569710;  
 RA Fukuchi K., Martin G.M., Deeb S.S.;  
 RT "Sequence of the protease inhibitor domain of the A4 amyloid protein  
 RT precursor of Mus domesticus.";  
 RL Nucleic Acids Res. 17:5396-5396(1989).  
 RN [9]  
 RP SEQUENCE OF 656-737 FROM N.A.  
 RC STRAIN=129/Sv;  
 RA Wragg M.A., Busfield F., Duff K., Korenblat K., Capecchi M.,  
 RA Loring J.F., Goate A.M.;  
 RT "Introduction of six mutations into the mouse genome using 'Hit and  
 RT Run' gene-targeting: introduction of familial Alzheimer's disease  
 RT mutations into the mouse amyloid precursor protein gene and  
 RT humanization of the A-beta fragment.";  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.  
 RN [10]  
 RP TISSUE SPECIFICITY OF ALTERNATIVE SPLICED FORMS.  
 RX MEDLINE=93287808; PubMed=8510506;  
 RA Sola C., Mengod G., Ghetti B., Palacios J.M., Triarhou L.C.;  
 RT "Regional distribution of the alternatively spliced isoforms of beta  
 RT APP RNA transcript in the brain of normal, heterozygous and  
 RT homozygous weaver mutant mice as revealed by in situ hybridization  
 RT histochemistry.";  
 RL Brain Res. Mol. Brain Res. 17:340-346(1993).  
 RN [11]  
 RP INTERACTION WITH KNS2.  
 RX MEDLINE=21010507; PubMed=11144355;  
 RA Kamal A., Stokin G.B., Yang Z., Xia C.-H., Goldstein L.S.;  
 RT "Axonal transport of amyloid precursor protein is mediated by direct  
 RT binding to the kinesin light chain subunit of kinesin-I.";  
 RL Neuron 28:449-459(2000).  
 RN [12]  
 RP C-TERMINAL PROTEIN-PROTEIN INTERACTIONS, AND MUTAGENESIS OF TYR-728;  
 RP THR-743; TYR-757; ASN-759 AND TYR-762.  
 RX MEDLINE=21408156; PubMed=11517249;  
 RA Matsuda S., Yasukawa T., Homma Y., Ito Y., Niikura T., Hiraki T.,  
 RA Hirai S., Ohno S., Kita Y., Kawasumi M., Kouyama K., Yamamoto T.,

RA Kyriakis J.M., Nishimoto I.;  
 RT "C-jun N-terminal kinase (JNK)-interacting protein-1b/islet-brain-1  
 RT scaffolds Alzheimer's amyloid precursor protein with JNK.";  
 RL J. Neurosci. 21:6597-6607(2001).  
 RN [13]  
 RP INTERACTION WITH MAPK8IP1, AND PHOSPHORYLATION.  
 RX MEDLINE=22028091; PubMed=11912189;  
 RA Taru H., Iijima K.-I., Hase M., Kirino Y., Yagi Y., Suzuki T.;  
 RT "Interaction of Alzheimer's beta-amyloid precursor family proteins  
 RT with scaffold proteins of the JNK signaling cascade.";  
 RL J. Biol. Chem. 277:20070-20078(2002).  
 RN [14]  
 RP INTERACTION OF CTF PEPTIDES WITH NUMB.  
 RX MEDLINE=22008109; PubMed=12011466;  
 RA Roncarati R., Sestan N., Scheinfeld M.H., Berechid B.E., Lopez P.A.,  
 RA Meucci O., McGlade J.C., Rakic P., D'Adamio L.;  
 RT "The gamma-secretase-generated intracellular domain of beta-amyloid  
 RT precursor protein binds Numb and inhibits Notch signaling.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:7102-7107(2002).  
 RN [15]  
 RP GAMMA-SECRETASE PROCESSING, AND INTERACTION WITH APBB1.  
 RX MEDLINE=21437805; PubMed=11553691;  
 RA Cupers P., Orlans I., Craessaerts K., Annaert W., De Strooper B.;  
 RT "The amyloid precursor protein (APP)-cytoplasmic fragment generated by  
 RT gamma-secretase is rapidly degraded but distributes partially in a  
 RT nuclear fraction of neurones in culture.";  
 RL J. Neurochem. 78:1168-1178(2001).  
 CC -!- FUNCTION: Functions as a cell surface receptor and performs  
 CC physiological functions on the surface of neurons relevant to  
 CC neurite growth, neuronal adhesion and axonogenesis. Involved in  
 CC cell mobility and transcription regulation through protein-protein  
 CC interactions. Can promote transcription activation through binding  
 CC to APBB1/Tip60 and inhibit Notch signaling through interaction  
 CC with Numb. Couples to apoptosis-inducing pathways such as those  
 CC mediated by G(0) and JIP. Inhibits G(0) alpha ATPase activity (By  
 CC similarity). Acts as a kinesin I membrane receptor, mediating the  
 CC axonal transport of beta-secretase and presenilin 1. May be  
 CC involved in copper homeostasis/oxidative stress through copper ion  
 CC reduction. Can regulate neurite outgrowth through binding to  
 CC components of the extracellular matrix such as heparin and  
 CC collagen I and IV (By similarity). The splice isoforms that  
 CC contain the BPTI domain possess protease inhibitor activity (By  
 CC similarity).  
 CC -!- FUNCTION: Beta-amyloid peptides are lipophilic metal chelators  
 CC with metal-reducing activity. Bind transient metals such as  
 CC copper, zinc and iron. Rat and mouse beta-amyloid peptides bind  
 CC only weakly transient metals and have little reducing activity due  
 CC to substitutions of transient metal chelating residues. Beta-APP42  
 CC may activate mononuclear phagocytes in the brain and elicit  
 CC inflammatory responses. Promotes both tau aggregation and TPK II-  
 CC mediated phosphorylation (By similarity).  
 CC -!- FUNCTION: The gamma-CTF peptides as well as the caspase-cleaved  
 CC peptides, including C31, are potent enhancers of neuronal  
 CC apoptosis.  
 CC -!- SUBUNIT: Binds, via its C-terminal, to the PID domain of several  
 CC cytoplasmic proteins, including APBB family members, the APBA  
 CC family, MAPK8IP1, SHC1, Numb and Dab1. Binding to Dab1 inhibits

CC its serine phosphorylation. Also interacts with GPCR-like protein  
 CC BPP, FPRL1, APPBP1, IBL, KNS2 (via its TPR domains), APPBP2 (via  
 CC BaSS) and DDB1 (By similarity). In vitro, it binds MAPT via the  
 CC MT-binding domains (By similarity). Associates with microtubules  
 CC in the presence of ATP and in a kinesin-dependent manner (By  
 CC similarity). Interacts, through a C-terminal domain, with GNAO1  
 CC (By similarity). Amyloid beta-42 binds CHRNA7 in hippocampal  
 CC neurons (By similarity). Beta-amyloid associates with HADH2 (By  
 CC similarity).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell surface  
 CC protein that rapidly becomes internalized via clathrin-coated  
 CC pits. During maturation, the immature APP (N-glycosylated in the  
 CC endoplasmic reticulum) moves to the Golgi complex where complete

Query Match 85.4%; Score 35; DB 1; Length 770;  
 Best Local Similarity 100.0%; Pred. No. 9.8;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVFFAED 7  
 |||||  
 Db 688 LVFFAED 694

# RESULT 11

## A4\_PIG

ID A4\_PIG STANDARD; PRT; 770 AA.  
 AC P79307; Q29023; Q9TUI0;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Amyloid beta A4 protein precursor (APP) (ABPP) (Alzheimer's disease  
 DE amyloid protein homolog) [Contains: Soluble APP-alpha (S-APP-alpha);  
 DE Soluble APP-beta (S-APP-beta); C99; Beta-amyloid protein 42 (Beta-  
 DE APP42); Beta-amyloid protein 40 (Beta-APP40); C83; P3(42); P3(40);  
 DE Gamma-CTF(59) (Gamma-secretase C-terminal fragment 59); Gamma-CTF(57)  
 DE (Gamma-secretase C-terminal fragment 57); Gamma-CTF(50) (Gamma-  
 DE secretase C-terminal fragment 50); C31].  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kimura A., Takahashi T.;  
 RT "Amyloid precursor protein 770.";  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE OF 1-136 FROM N.A.  
 RC TISSUE=Small intestine;  
 RA Winteroe A.K., Fredholm M.;  
 RT "Evaluation and characterization of a porcine small intestine cDNA  
 RT library.";  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE OF 667-723 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=92017079; PubMed=1656157;

RA Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;  
 RT "Conservation of the sequence of the Alzheimer's disease amyloid  
 RT peptide in dog, polar bear and five other mammals by cross-species  
 RT polymerase chain reaction analysis.";  
 RL Brain Res. Mol. Brain Res. 10:299-305(1991).

CC -!- FUNCTION: Functions as a cell surface receptor and performs  
 CC physiological functions on the surface of neurons relevant to  
 CC neurite growth, neuronal adhesion and axonogenesis. Involved in  
 CC cell mobility and transcription regulation through protein-protein  
 CC interactions (By similarity). Can promote transcription activation  
 CC through binding to APBB1/Tip60 and inhibit Notch signaling through  
 CC interaction with Numb (By similarity). Couples to apoptosis-  
 CC inducing pathways such as those mediated by G(O) and JIP (By  
 CC similarity). Inhibits G(O) alpha ATPase activity (By similarity).  
 CC Acts as a kinesin I membrane receptor, mediating the axonal  
 CC transport of beta-secretase and presenilin 1 (By similarity). May  
 CC be involved in copper homeostasis/oxidative stress through copper  
 CC ion reduction (By similarity). In vitro, copper-metallated APP  
 CC induces neuronal death directly or is potentiated through Cu(II)-  
 CC mediated low-density lipoprotein oxidation (By similarity). Can  
 CC regulate neurite outgrowth through binding to components of the  
 CC extracellular matrix such as heparin and collagen I and IV (By  
 CC similarity).

CC -!- FUNCTION: Beta-amyloid peptides are lipophilic metal chelators  
 CC with metal-reducing activity. Bind transient metals such as  
 CC copper, zinc and iron (By similarity).

CC -!- FUNCTION: The gamma-CTF peptides as well as the caspase-cleaved  
 CC peptides, including C31, are potent enhancers of neuronal  
 CC apoptosis (By similarity).

CC -!- SUBUNIT: Binds, via its C-terminal, to the PID domain of several  
 CC cytoplasmic proteins, including APBB family members, the APBA  
 CC family, MAPK8IP1, and SHC1, Numb and Dab1 (By similarity). Binding  
 CC to Dab1 inhibits its serine phosphorylation (By similarity). Also  
 CC interacts with GPCR-like protein BPP, FPRL1, APPBP1, IB1, KNS2  
 CC (via its TPR domains) (By similarity), APPBP2 (via BaSS) and DDB1.  
 CC In vitro, it binds MAPT via the MT-binding domains (By  
 CC similarity). Associates with microtubules in the presence of ATP  
 CC and in a kinesin-dependent manner (By similarity).

CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell surface  
 CC protein that rapidly becomes internalized via clathrin-coated  
 CC pits. During maturation, the immature APP (N-glycosylated in the  
 CC endoplasmic reticulum) moves to the Golgi complex where complete  
 CC maturation occurs (O-glycosylated and sulfated). After alpha-  
 CC secretase cleavage, soluble APP is released into the extracellular  
 CC space and the C-terminal is internalized to endosomes and  
 CC lysosomes. Some APP accumulates in secretory transport vesicles  
 CC leaving the late Golgi compartment and returns to the cell  
 CC surface. Gamma-CTF(59) peptide is located to both the cytoplasm  
 CC and nuclei of neurons (By similarity).

CC -!- DOMAIN: The basolateral sorting signal (BaSS) is required for  
 CC sorting of membrane proteins to the basolateral surface of  
 CC epithelial cells (By similarity).

CC -!- DOMAIN: The NPXY sequence motif found in many tyrosine-  
 CC phosphorylated proteins is required for the specific binding of  
 CC the PID domain. However additional amino acids either N- or C-  
 CC terminal to the NPXY motif are often required for complete  
 CC interaction. The PID domain-containing proteins which bind APP



require the YENPTY motif for full interaction. These interactions are independent of phosphorylation on the terminal tyrosine residue. The NPXY site is also involved in clathrin-mediated endocytosis (By similarity).

-!- PTM: Proteolytically processed under normal cellular conditions. Cleavage by alpha-secretase or alternatively by beta-secretase leads to generation and extracellular release of soluble APP peptides, S-APP-alpha and S-APP-beta, respectively, and the retention of corresponding membrane-anchored C-terminal fragments, C83 and C99. Subsequent processing of C83 by gamma-secretase yields P3 peptides. This is the major secretory pathway and is nonamyloidogenic. Alternatively, presenilin/nicastrin-mediated gamma-secretase processing of C99 releases the amyloid beta proteins, amyloid-beta 40 (Abeta40) and amyloid-beta 42 (Abeta42), major components of amyloid plaques, and the cytotoxic C-terminal fragments, gamma-CTF(50), gamma-CTF(57) and gamma-CTF(59) (By similarity).

-!- PTM: Proteolytically cleaved by caspases during neuronal apoptosis (By similarity). Cleavage at Asp-739 by either caspase-3, -8 or -9 results in the production of the neurotoxic C31 peptide and the increased production of beta-amyloid peptides (By similarity).

-!- PTM: N- and O-glycosylated (By similarity).

-!- PTM: Phosphorylation in the C-terminal on tyrosine, threonine and serine residues is neuron-specific. Phosphorylation can affect APP processing, neuronal differentiation and interaction with other proteins (By similarity).

-!- PTM: Extracellular binding and reduction of copper, results in a corresponding oxidation of Cys-144 and Cys-158, and the formation of a disulfide bond (By similarity).

-!- MISCELLANEOUS: Chelation of metal ions, notably copper, iron and zinc, can induce histidine-bridging between beta-amyloid molecules resulting in beta-amyloid-metal aggregates (By similarity). Extracellular zinc-binding increases binding of heparin to APP and inhibits collagen-binding (By similarity).

-!- SIMILARITY: Belongs to the APP family.

-!- SIMILARITY: Contains 1 BPTI/Kunitz inhibitor domain.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

DR EMBL; AB032550; BAA84580.1; -.  
 DR EMBL; Z84022; CAB06313.1; -.  
 DR EMBL; X56127; CAA39592.1; -.  
 DR HSSP; P05067; 1AAP.  
 DR InterPro; IPR008155; A4\_APP.  
 DR InterPro; IPR008154; A4\_extra.  
 DR InterPro; IPR002223; Kunitz\_BPTI.  
 DR Pfam; PF02177; A4\_EXTRA; 1.  
 DR PRINTS; PR00203; AMYLOIDA4.  
 DR PRINTS; PR00759; BASICPTASE.  
 DR ProDom; PD000222; Kunitz\_BPTI; 1.  
 DR SMART; SM00006; A4\_EXTRA; 1.

DR SMART; SM00131; KU; 1.  
 DR PROSITE; PS00319; A4\_EXTRA; 1.  
 DR PROSITE; PS00320; A4\_INTRA; 1.  
 DR PROSITE; PS00280; BPTI\_KUNITZ\_1; 1.  
 DR PROSITE; PS50279; BPTI\_KUNITZ\_2; 1.  
 KW Apoptosis; Endocytosis; Cell adhesion; Serine protease inhibitor;  
 KW Coated pits; Neurone; Heparin-binding; Metal-binding; Copper; Iron;  
 KW Zinc; Signal; Transmembrane; Glycoprotein; Phosphorylation;  
 KW Amyloid.

FT	SIGNAL	1	17	BY SIMILARITY.
FT	CHAIN	18	770	AMYLOID BETA A4 PROTEIN.
FT	CHAIN	18	687	SOLUBLE APP-ALPHA (POTENTIAL).
FT	CHAIN	18	671	SOLUBLE APP-BETA (POTENTIAL).
FT	CHAIN	672	770	C99 (BY SIMILARITY).
FT	CHAIN	672	713	BETA-AMYLOID PROTEIN 42 (BY SIMILARITY).
FT	CHAIN	672	711	BETA-AMYLOID PROTEIN 40 (BY SIMILARITY).
FT	CHAIN	688	770	C83 (BY SIMILARITY).
FT	CHAIN	688	713	P3(42) (BY SIMILARITY).
FT	CHAIN	688	711	P3(40) (BY SIMILARITY).
FT	CHAIN	712	770	GAMMA-CTF(59).
FT	CHAIN	714	770	GAMMA-CTF(57).
FT	CHAIN	721	770	GAMMA-CTF(50) (BY SIMILARITY).
FT	CHAIN	740	770	C31 (DURING APOPTOSIS) (BY SIMILARITY).
FT	DOMAIN	18	699	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	700	723	POTENTIAL.
FT	DOMAIN	724	770	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	96	110	HEPARIN-BINDING (BY SIMILARITY).
FT	DOMAIN	135	155	COPPER-BINDING (BY SIMILARITY).
FT	DOMAIN	181	188	ZINC-BINDING (BY SIMILARITY).
FT	DOMAIN	291	341	BPTI/KUNITZ INHIBITOR.
FT	DOMAIN	391	423	HEPARIN-BINDING (BY SIMILARITY).
FT	DOMAIN	491	522	HEPARIN-BINDING (BY SIMILARITY).
FT	DOMAIN	523	540	COLLAGEN-BINDING (BY SIMILARITY).
FT	DOMAIN	732	751	INTERACTION WITH G(O)-ALPHA (BY SIMILARITY).
FT				
FT	DOMAIN	230	260	ASP/GLU-RICH (ACIDIC).
FT	DOMAIN	274	280	POLY-THR.
FT	SITE	144	144	REQUIRED FOR COPPER(II) REDUCTION (BY SIMILARITY).
FT				
FT	ACT_SITE	301	302	REACTIVE BOND (BY SIMILARITY).
FT	SITE	671	672	CLEAVAGE (BY BETA-SECRETASE) (BY SIMILARITY).
FT				
FT	SITE	672	673	CLEAVAGE (BY CASPASE-6) (BY SIMILARITY).
FT	SITE	687	688	CLEAVAGE (BY ALPHA-SECRETASE) (BY SIMILARITY).
FT				
FT	SITE	704	704	IMPLICATED IN FREE RADICAL PROPAGATION (BY SIMILARITY).
FT				
FT	SITE	706	706	INVOLVED IN OXIDATIVE REACTIONS (BY SIMILARITY).
FT				
FT	SITE	711	712	CLEAVAGE (BY GAMMA-SECRETASE; SITE 1) (BY SIMILARITY).
FT				
FT	SITE	713	714	CLEAVAGE (BY GAMMA-SECRETASE; SITE 2) (BY SIMILARITY).
FT				
FT	SITE	720	721	CLEAVAGE (BY GAMMA-SECRETASE; SITE 3)

Query Match 85.4%; Score 35; DB 1; Length 770;  
 Best Local Similarity 100.0%; Pred. No. 9.8;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
|||||||  
Db 688 LVFFAED 694

RESULT 12

A4\_RAT

ID A4\_RAT STANDARD; PRT; 770 AA.

AC P08592;

DT 01-AUG-1988 (Rel. 08, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Amyloid beta A4 protein precursor (APP) (Alzheimer's disease amyloid  
DE protein homolog) (Amyloidogenic glycoprotein) (AG) [Contains: Soluble  
DE APP-alpha (S-APP-alpha); Soluble APP-beta (S-APP-beta); C99; Beta-  
DE amyloid protein 42 (Beta-APP42); Beta-amyloid protein 40 (Beta-APP40);  
DE C83; P3(42); P3(40); Gamma-CTF(59) (Gamma-secretase C-terminal  
DE fragment 59); Gamma-CTF(57) (Gamma-secretase C-terminal fragment 57);  
DE Gamma-CTF(50) (Gamma-secretase C-terminal fragment 50); C31].

GN APP.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM APP695).

RC TISSUE=Brain;

RX MEDLINE=88312583; PubMed=2900758;

RA Shivers B.D., Hilbich C., Multhaup G., Salbaum J.M., Beyreuther K.,

RA Seeburg P.H.;

RT "Alzheimer's disease amyloidogenic glycoprotein: expression pattern

RT in rat brain suggests a role in cell contact.";

RL EMBO J. 7:1365-1370(1988).

RN [2]

RP SEQUENCE OF 289-364 FROM N.A.

RC TISSUE=Liver;

RX MEDLINE=89183625; PubMed=2648331;

RA Kang J., Mueller-Hill B.;

RT "The sequence of the two extra exons in rat preA4.";

RL Nucleic Acids Res. 17:2130-2130(1989).

RN [3]

RP SEQUENCE OF 720-730, AND MASS SPECTROMETRY.

RX MEDLINE=21443797; PubMed=11483588;

RA Gu Y., Misonou H., Sato T., Dohmae N., Takio K., Ihara Y.;

RT "Distinct intramembrane cleavage of the beta-amyloid precursor protein

RT family resembling gamma-secretase-like cleavage of Notch.";

RL J. Biol. Chem. 276:35235-35238(2001).

RN [4]

RP ALTERNATIVE SPLICING.

RX MEDLINE=96187032; PubMed=8624099;

RA Sandbrink R., Masters C.L., Beyreuther K.;

RT "APP gene family. Alternative splicing generates functionally related  
RT isoforms.";

RL Ann. N.Y. Acad. Sci. 777:281-287(1996).

RN [5]

RP TISSUE SPECIFICITY OF APPICAN.  
RX MEDLINE=95263526; PubMed=7744833;  
RA Shioi J., Pangalos M.N., Ripellino J.A., Vassilacopoulou D.,  
RA Mytilineou C., Margolis R.U., Robakis N.K.;  
RT "The Alzheimer amyloid precursor proteoglycan (appican) is present in  
RT brain and is produced by astrocytes but not by neurons in primary  
RT neural cultures.";  
RL J. Biol. Chem. 270:11839-11844(1995).  
RN [6]

RP TISSUE SPECIFICITY OF ISOFORMS.  
RX MEDLINE=97150061; PubMed=8996834;  
RA Sandbrink R., Monning U., Masters C.L., Beyreuther K.;  
RT "Expression of the APP gene family in brain cells, brain development  
RT and aging.";  
RL Gerontology 43:119-131(1997).  
RN [7]

RP INTERACTION WITH DDB1, AND MUTAGENESIS OF TYR-757; ASN-759 AND  
RP TYR-762.  
RX MEDLINE=99127916; PubMed=9930726;  
RA Watanabe T., Sukegawa J., Tomita S., Iijima K.-I., Oguchi S.,  
RA Suzuki T., Nairn A.C., Greengard P.;  
RT "A 127-kDa protein (UV-DDB) binds to the cytoplasmic domain of the  
RT Alzheimer's amyloid precursor protein.";  
RL J. Neurochem. 72:549-556(1999).  
RN [8]

RP INTERACTION WITH GNAO1, AND MUTAGENESIS OF 732-HIS-HIS-733.  
RX MEDLINE=99162676; PubMed=10024358;  
RA Brouillet E., Trembleau A., Galanaud D., Volovitch M., Bouillot C.,  
RA Valenza C., Prochiantz A., Allinquant B.;  
RT "The amyloid precursor protein interacts with Go heterotrimeric  
RT protein within a cell compartment specialized in signal  
RT transduction.";  
RL J. Neurosci. 19:1717-1727(1999).  
RN [9]

RP CHARACTERISTICS OF APPICAN, AND MUTAGENESIS OF SER-656.  
RX MEDLINE=95256193; PubMed=7737970;  
RA Pangalos M.N., Efthimiopoulos S., Shioi J., Robakis N.K.;  
RT "The chondroitin sulfate attachment site of appican is formed by  
RT splicing out exon 15 of the amyloid precursor gene.";  
RL J. Biol. Chem. 270:10388-10391(1995).  
RN [10]

RP BETA-AMYLOID METAL-BINDING.  
RX MEDLINE=99316162; PubMed=10386999;  
RA Huang X., Atwood C.S., Hartshorn M.A., Multhaup G., Goldstein L.E.,  
RA Scarpa R.C., Cuajungco M.P., Gray D.N., Lim J., Moir R.D., Tanzi R.E.,  
RA Bush A.I.;  
RT "The A beta peptide of Alzheimer's disease directly produces hydrogen  
RT peroxide through metal ion reduction.";  
RL Biochemistry 38:7609-7616(1999).  
RN [11]

RP BETA-AMYLOID ZINC BINDING.  
RX MEDLINE=99343552; PubMed=10413512;  
RA Liu S.T., Howlett G., Barrow C.J.;  
RT "Histidine-13 is a crucial residue in the zinc ion-induced aggregation  
RT of the A beta peptide of Alzheimer's disease.";  
RL Biochemistry 38:9373-9378(1999).  
RN [12]

RP IMPORTANCE OF GLY-704 IN FREE RADICAL PROPAGATION, AND MUTAGENESIS OF  
 RP GLY-704.  
 RX MEDLINE=21956095; PubMed=11959460;  
 RA Kanski J., Varadarajan S., Aksenova M., Butterfield D.A.;  
 RT "Role of glycine-33 and methionine-35 in Alzheimer's amyloid beta-  
 RT peptide 1-42-associated oxidative stress and neurotoxicity.";  
 RL Biochim. Biophys. Acta 1586:190-198(2001).  
 RN [13]  
 RP PHOSPHORYLATION.  
 RX MEDLINE=97239592; PubMed=9085254;  
 RA Oishi M., Nairn A.C., Czernik A.J., Lim G.S., Isohara T., Gandy S.E.,  
 RA Greengard P., Suzuki T.;  
 RT "The cytoplasmic domain of Alzheimer's amyloid precursor protein is  
 RT phosphorylated at Thr654, Ser655, and Thr668 in adult rat brain and  
 RT cultured cells.";  
 RL Mol. Med. 3:111-123(1997).  
 RN [14]  
 RP PHOSPHORYLATION ON SER-730.  
 RX MEDLINE=99262094; PubMed=10329382;  
 RA Isohara T., Horiuchi A., Watanabe T., Ando K., Czernik A.J., Uno I.,  
 RA Greengard P., Nairn A.C., Suzuki T.;  
 RT "Phosphorylation of the cytoplasmic domain of Alzheimer's beta-amyloid  
 RT precursor protein at Ser655 by a novel protein kinase.";  
 RL Biochem. Biophys. Res. Commun. 258:300-305(1999).  
 RN [15]  
 RP PHOSPHORYLATION, INDUCTION, SUBCELLULAR LOCATION, AND MUTAGENESIS OF  
 RP THR-743.  
 RX MEDLINE=99274744; PubMed=10341243;  
 RA Ando K., Oishi M., Takeda S., Iijima K.-I., Isohara T., Nairn A.C.,  
 RA Kirino Y., Greengard P., Suzuki T.;  
 RT "Role of phosphorylation of Alzheimer's amyloid precursor protein  
 RT during neuronal differentiation.";  
 RL J. Neurosci. 19:4421-4427(1999).  
 RN [16]  
 RP PHOSPHORYLATION ON THR-743.  
 RX MEDLINE=20396183; PubMed=10936190;  
 RA Iijima K.-I., Ando K., Takeda S., Satoh Y., Seki T., Itohara S.,  
 RA Greengard P., Kirino Y., Nairn A.C., Suzuki T.;  
 RT "Neuron-specific phosphorylation of Alzheimer's beta-amyloid precursor  
 RT protein by cyclin-dependent kinase 5.";  
 RL J. Neurochem. 75:1085-1091(2000).  
 RN [17]  
 RP CARBOHYDRATE STRUCTURE OF APPICAN.  
 RX MEDLINE=21463085; PubMed=11479316;  
 RA Tsuchida K., Shioi J., Yamada S., Boghosian G., Wu A., Cai H.,  
 RA Sugahara K., Robakis N.K.;  
 RT "Appican, the proteoglycan form of the amyloid precursor protein,  
 RT contains chondroitin sulfate E in the repeating disaccharide region  
 RT and 4-O-sulfated galactose in the linkage region.";  
 RL J. Biol. Chem. 276:37155-37160(2001).  
 CC -!- FUNCTION: Functions as a cell surface receptor and performs  
 CC physiological functions on the surface of neurons relevant to  
 CC neurite growth, neuronal adhesion and axonogenesis. Involved in  
 CC cell mobility and transcription regulation through protein-protein  
 CC interactions (By similarity). Can promote transcription activation  
 CC through binding to APBB1/Tip60 and inhibit Notch signaling through  
 CC interaction with Numb (By similarity). Couples to apoptosis-

CC inducing pathways such as those mediated by G(0) and JIP. Inhibits  
 CC G(0) alpha ATPase activity. Acts as a kinesin I membrane receptor,  
 CC mediating the axonal transport of beta-secretase and presenilin 1  
 CC (By similarity). May be involved in copper homeostasis/oxidative  
 CC stress through copper ion reduction. Can regulate neurite  
 CC outgrowth through binding to components of the extracellular  
 CC matrix such as heparin and collagen I and IV (By similarity). The  
 CC splice isoforms that contain the BPTI domain possess protease  
 CC inhibitor activity (By similarity).  
 CC -!- FUNCTION: Beta-amyloid peptides are lipophilic metal chelators  
 CC with metal-reducing activity. Bind transient metals such as  
 CC copper, zinc and iron. Rat and mouse beta-amyloid peptides bind  
 CC only weakly transient metals and have little reducing activity due  
 CC to substitutions of transient metal chelating residues. Beta-APP42  
 CC may activate mononuclear phagocytes in the brain and elicit  
 CC inflammatory responses. Promotes both tau aggregation and TPK II-  
 CC mediated phosphorylation (By similarity).  
 CC -!- FUNCTION: Appicans elicit adhesion of neural cells to the  
 CC extracellular matrix and may regulate neurite outgrowth in the  
 CC brain.  
 CC -!- FUNCTION: The gamma-CTF peptides as well as the caspase-cleaved  
 CC peptides, including C31, are potent enhancers of neuronal  
 CC apoptosis (By similarity).  
 CC -!- SUBUNIT: Binds, via its C-terminal, to the PID domain of several  
 CC cytoplasmic proteins, including APBB family members, the APBA  
 CC family, MAPK8IP1, SHC1 and Numb and Dab1 (By similarity). Binding  
 CC to Dab1 inhibits its serine phosphorylation (By similarity). Also  
 CC interacts with GPCR-like protein BPP, FPRL1, APPBP1, IB1, KNS2  
 CC (via its TPR domains), APPBP2 (via BaSS) (By similarity) and DDB1.  
 CC In vitro, it binds MAPT via the MT-binding domains (By  
 CC similarity). Associates with microtubules in the presence of ATP  
 CC and in a kinesin-dependent manner (By similarity). Interacts,  
 CC through a C-terminal domain, with GNAO1. Amyloid beta-42 binds  
 CC CHRNA7 in hippocampal neurons (By similarity). Beta-amyloid  
 CC associates with HADH2 (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell surface  
 CC protein that rapidly becomes internalized via clathrin-coated  
 CC pits. During maturation, the immature APP (N-glycosylated in the

Query Match 85.4%; Score 35; DB 1; Length 770;  
 Best Local Similarity 100.0%; Pred. No. 9.8;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
 |||||  
 Db 688 LVFFAED 694

#### RESULT 13

##### A4\_TETFL

ID A4\_TETFL STANDARD; PRT; 780 AA.  
 AC O73683;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Alzheimer's disease amyloid A4 protein homolog precursor [Contains:  
 DE Beta-amyloid protein (Beta-APP) (A-beta)].

GN APP.  
 OS Tetraodon fluviatilis (Puffer fish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 OC Tetraodontoidea; Tetraodontidae; Tetraodon.  
 OX NCBI\_TaxID=47145;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98252138; PubMed=9599080;  
 RA Villard L., Tassone F., Crnogorac-Jurcevic T., Clancy K., Gardiner K.;  
 RT "Analysis of pufferfish homologues of the AT-rich human APP gene."  
 RL Gene 210:17-24(1998).  
 CC -!- FUNCTION: Functional neuronal receptor which couples to  
 CC intracellular signaling pathway through the GTP-binding protein  
 CC G(0) (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- SIMILARITY: Belongs to the APP family.  
 CC -!- SIMILARITY: Contains 1 BPTI/Kunitz inhibitor domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF018165; AAC41275.1; -.  
 DR HSSP; P05067; 1HZ3.  
 DR InterPro; IPR008155; A4\_APP.  
 DR InterPro; IPR008154; A4\_extra.  
 DR InterPro; IPR001255; Beta-APP.  
 DR InterPro; IPR002223; Kunitz\_BPTI.  
 DR Pfam; PF02177; A4\_EXTRA; 1.  
 DR Pfam; PF03494; Beta-APP; 1.  
 DR Pfam; PF00014; Kunitz\_BPTI; 1.  
 DR PRINTS; PR00203; AMYLOIDA4.  
 DR PRINTS; PR00759; BASICPTASE.  
 DR ProDom; PD000222; Kunitz\_BPTI; 1.  
 DR SMART; SM00006; A4\_EXTRA; 1.  
 DR SMART; SM00131; KU; 1.  
 DR PROSITE; PS00319; A4\_EXTRA; 1.  
 DR PROSITE; PS00320; A4\_INTRA; 1.  
 DR PROSITE; PS00280; BPTI\_KUNITZ\_1; FALSE\_NEG.  
 DR PROSITE; PS50279; BPTI\_KUNITZ\_2; 1.  
 KW Glycoprotein; Amyloid; Neurone; Transmembrane; Signal;  
 KW Serine protease inhibitor.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT CHAIN 19 780 ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN  
 FT HOMOLOG.  
 FT CHAIN 682 724 BETA-AMYLOID PROTEIN (POTENTIAL).  
 FT DOMAIN 19 711 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 712 732 POTENTIAL.  
 FT DOMAIN 733 780 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 323 382 BPTI/KUNITZ INHIBITOR.  
 FT SITE 769 772 CLATHRIN-BINDING (BY SIMILARITY).

FT DISULFID 327 378 BY SIMILARITY.  
 FT DISULFID 336 361 BY SIMILARITY.  
 FT CARBOHYD 560 560 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 780 AA; 88238 MW; 60071BE94520191D CRC64;

Query Match 85.4%; Score 35; DB 1; Length 780;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
 |||||  
 Db 698 LVFFAED 704

#### RESULT 14

##### PE23\_SHEEP

ID PE23\_SHEEP STANDARD; PRT; 89 AA.  
 AC Q28550;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Prostaglandin E2 receptor, EP3 subtype (Prostanoid EP3 receptor) (PGE  
 DE receptor, EP3 subtype) (Fragment).  
 GN PTGER3.  
 OS Ovis aries (Sheep).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
 OC Bovidae; Caprinae; Ovis.  
 OX NCBI\_TaxID=9940;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney outer medulla;  
 RX MEDLINE=98287159; PubMed=9625477;  
 RA Audicana L., Aughey E., O'Shaughnessy P.J.;  
 RT "Sensitivity of the early luteal phase ovine cervix to prostaglandin  
 RT E2 (PGE2) and expression of EP3 receptor mRNA."  
 RL Res. Vet. Sci. 64:177-179(1998).  
 CC -!- FUNCTION: Receptor for prostaglandin E2 (PGE2); the EP3 receptor  
 CC may be involved in inhibition of gastric acid secretion,  
 CC modulation of neurotransmitter release in central and peripheral  
 CC neurons, inhibition of sodium and water reabsorption in kidney  
 CC tubulus and contraction in uterine smooth muscle. The activity of  
 CC this receptor can couple to both the inhibition of adenylate  
 CC cyclase mediated by G(i) proteins, and to an elevation of  
 CC intracellular calcium (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U37148; AAB81195.1; -.



DR InterPro; IPR000276; GPCR\_Rhodpsn.  
 DR PROSITE; PS00237; G\_PROTEIN\_RECEP\_F1\_1; FALSE\_NEG.  
 KW G-protein coupled receptor; Transmembrane; Glycoprotein.  
 FT NON\_TER 1 1  
 FT TRANSMEM <1 18 4 (POTENTIAL).  
 FT DOMAIN 19 48 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 49 74 5 (POTENTIAL).  
 FT DOMAIN 75 89 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT NON\_TER 89 89  
 SQ SEQUENCE 89 AA; 9364 MW; EDAD27E127B0A428 CRC64;

Query Match 78.0%; Score 32; DB 1; Length 89;  
 Best Local Similarity 85.7%; Pred. No. 5.1;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
 |||| ||  
 Db 49 VFFASDF 55

#### RESULT 15

A4\_FUGRU

ID A4\_FUGRU STANDARD; PRT; 737 AA.

AC O93279;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Alzheimer's disease amyloid A4 protein homolog precursor [Contains:

DE Beta-amyloid protein (Beta-APP) (A-beta)].

GN APP.

OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;

OC Tetradontoidea; Tetraodontidae; Takifugu.

OX NCBI\_TaxID=31033;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98252138; PubMed=9599080;

RA Villard L., Tassone F., Crnogorac-Jurcevic T., Clancy K., Gardiner K.;

RT "Analysis of pufferfish homologues of the AT-rich human APP gene.";

RL Gene 210:17-24(1998).

CC -!- FUNCTION: Functional neuronal receptor which couples to  
 CC intracellular signaling pathway through the GTP-binding protein  
 CC G(O) (By similarity).

CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

CC -!- SIMILARITY: Belongs to the APP family.

CC -!- SIMILARITY: Contains 1 BPTI/Kunitz inhibitor domain.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
DR EMBL; AF090120; AAD13392.1; -.
DR HSSP; P05067; 1HZ3.
DR InterPro; IPR008155; A4_APP.
DR InterPro; IPR008154; A4_extra.
DR InterPro; IPR001255; Beta-APP.
DR InterPro; IPR002223; Kunitz_BPTI.
DR Pfam; PF02177; A4_EXTRA; 1.
DR Pfam; PF03494; Beta-APP; 1.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR PRINTS; PR00203; AMYLOIDA4.
DR PRINTS; PR00759; BASICPTASE.
DR ProDom; PD000222; Kunitz_BPTI; 1.
DR SMART; SM00006; A4_EXTRA; 1.
DR SMART; SM00131; KU; 1.
DR PROSITE; PS00319; A4_EXTRA; FALSE_NEG.
DR PROSITE; PS00320; A4_INTRA; 1.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS50279; BPTI_KUNITZ_2; 1.
KW Glycoprotein; Amyloid; Neurone; Transmembrane; Signal;
KW Serine protease inhibitor.
FT SIGNAL      1      18      POTENTIAL.
FT CHAIN       19     737     ALZHEIMER'S DISEASE AMYLOID A4
FT                                     PROTEIN HOMOLOG.
FT CHAIN       639     681     BETA-AMYLOID PROTEIN (POTENTIAL).
FT DOMAIN      19     668     EXTRACELLULAR (POTENTIAL).
FT TRANSMEM    669     689     POTENTIAL.
FT DOMAIN      690     737     CYTOPLASMIC (POTENTIAL).
FT DOMAIN      286     344     BPTI/KUNITZ INHIBITOR.
FT SITE        726     729     CLATHRIN-BINDING (BY SIMILARITY).
FT ACT_SITE    300     301     REACTIVE BOND.
FT DISULFID    290     340     BY SIMILARITY.
FT DISULFID    299     323     BY SIMILARITY.
FT DISULFID    315     336     BY SIMILARITY.
FT CARBOHYD    522     522     N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE    737 AA;  82856 MW;  6FAD01E2E3B2B7E2 CRC64;

```

```

Query Match          78.0%;  Score 32;  DB 1;  Length 737;
Best Local Similarity 85.7%;  Pred. No. 42;
Matches      6;  Conservative      1;  Mismatches      0;  Indels      0;  Gaps      0;

```

```

QY      1 LVFFAED 7
        |||||:|
Db      655 LVFFADD 661

```

```

RESULT 16
Y691_CHLTR
ID Y691_CHLTR      STANDARD;      PRT;  224 AA.
AC 084697;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein CT691.
GN CT691.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.

```

OX NCBI\_TaxID=813;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=D/UW-3/Cx;  
 RX MEDLINE=99000809; PubMed=9784136;  
 RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
 RA Davis R.W.;  
 RT "Genome sequence of an obligate intracellular pathogen of humans:  
 RT Chlamydia trachomatis.";  
 RL Science 282:754-759(1998).  
 CC -!- SIMILARITY: Belongs to the UPF0111 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE001339; AAC68286.1; -.  
 DR PIR; G71483; G71483.  
 DR InterPro; IPR002727; DUF47.  
 DR Pfam; PF01865; PhoU\_div; 1.  
 DR TIGRFAMs; TIGR00153; TIGR00153; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 224 AA; 25176 MW; 661EEB640DD1401E CRC64;

Query Match 75.6%; Score 31; DB 1; Length 224;  
 Best Local Similarity 71.4%; Pred. No. 21;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 VFFAEDF 8  
 |||::||  
 Db 180 VFFSDDF 186

#### RESULT 17

##### UPK\_CORST

ID UPK\_CORST STANDARD; PRT; 281 AA.  
 AC Q9FB58;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Putative undecaprenol kinase (EC 2.7.1.66) (Bacitracin resistance  
 DE protein).  
 GN UPK OR BACA.  
 OS Corynebacterium striatum.  
 OG Plasmid pTP10.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
 OX NCBI\_TaxID=43770;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=M82B;  
 RX MEDLINE=20194806; PubMed=10732668;

RA Tauch A., Krieft S., Kalinowski J., Puehler A.;  
 RT "The 51,409-bp R-plasmid pTP10 from the multiresistant clinical  
 RT isolate *Corynebacterium striatum* M82B is composed of DNA segments  
 RT initially identified in soil bacteria and in plant, animal, and human  
 RT pathogens.";  
 RL Mol. Gen. Genet. 263:1-11(2000).  
 CC -!- FUNCTION: Probably phosphorylates undecaprenol to undecaprenyl  
 CC phosphate. Confers resistance to bacitracin (By similarity).  
 CC -!- CATALYTIC ACTIVITY: ATP + undecaprenol = ADP + undecaprenyl  
 CC phosphate.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -!- MISCELLANEOUS: Bacitracin is thought to be involved in inhibition  
 CC of peptidoglycan synthesis by sequestering undecaprenyl  
 CC diphosphate reducing the pool of lipid carrier available.  
 CC -!- SIMILARITY: Belongs to the upk family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF024666; AAG03365.1; -.  
 DR HAMAP; MF\_01006; -; 1.  
 DR InterPro; IPR003824; BacA.  
 DR Pfam; PF02673; BacA; 1.  
 DR TIGRFAMs; TIGR00753; undec\_kin\_bacA; 1.  
 KW Transferase; Kinase; Antibiotic resistance; Transmembrane; Plasmid.  
 FT TRANSMEM 13 32 POTENTIAL.  
 FT TRANSMEM 52 71 POTENTIAL.  
 FT TRANSMEM 98 115 POTENTIAL.  
 FT TRANSMEM 125 142 POTENTIAL.  
 FT TRANSMEM 198 220 POTENTIAL.  
 FT TRANSMEM 230 252 POTENTIAL.  
 FT TRANSMEM 259 278 POTENTIAL.  
 SQ SEQUENCE 281 AA; 30512 MW; F82700075DB27EA9 CRC64;

Query Match 75.6%; Score 31; DB 1; Length 281;  
 Best Local Similarity 85.7%; Pred. No. 26;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
 ||||:|  
 Db 67 LVFFAKD 73

RESULT 18  
 YWBI\_BACSU  
 ID YWBI\_BACSU STANDARD; PRT; 301 AA.  
 AC P39592;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Putative HTH-type transcriptional regulator ywBI.  
 GN YWBI OR IPA-24D OR BSU38310.

OS *Bacillus subtilis*.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; *Bacillus*.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=95020537; PubMed=7934828;  
 RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,  
 RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,  
 RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,  
 RA Rapoport G., Danchin A.;  
 RT "Bacillus subtilis genome project: cloning and sequencing of the 97  
 RT kb region from 325 degrees to 333 degrees.";  
 RL Mol. Microbiol. 10:371-384(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*.";  
 RL Nature 390:249-256(1997).  
 CC -!- SIMILARITY: Contains 1 HTH lysR-type DNA-binding domain.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X73124; CAA51580.1; -.

DR EMBL; Z99123; CAB15857.1; -.

DR PIR; S39679; S39679.

DR SubtiList; BG10570; ywbI.

DR InterPro; IPR000847; HTH\_LysR.

DR InterPro; IPR005119; LysR\_subst.

DR Pfam; PF00126; HTH\_1; 1.

DR Pfam; PF03466; LysR\_substrate; 1.

DR PRINTS; PR00039; HTHLYSR.

DR PROSITE; PS50931; HTH\_LYSR; 1.

KW Hypothetical protein; Transcription regulation; DNA-binding;

KW Complete proteome.

FT DOMAIN 1 58 HTH\_LYSR-TYPE.

FT DNA\_BIND 18 37 H-T-H MOTIF (POTENTIAL).

SQ SEQUENCE 301 AA; 34677 MW; 63B06C0A7D3D30B9 CRC64;

Query Match 75.6%; Score 31; DB 1; Length 301;

Best Local Similarity 71.4%; Pred. No. 28;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8

:|| |||

Db 192 IFFREDF 198

#### RESULT 19

##### MM14\_PIG

ID MM14\_PIG STANDARD; PRT; 580 AA.

AC Q9XT90;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Matrix metalloproteinase-14 precursor (EC 3.4.24.80) (MMP-14)

DE (Membrane-type matrix metalloproteinase 1) (MT-MMP 1) (MTMMP1)

DE (Membrane-type-1 matrix metalloproteinase) (MT1-MMP) (MT1MMP).

GN MMP14.

OS Sus scrofa (Pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

OX NCBI\_TaxID=9823;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99095929; PubMed=9881602;

RA Caron C., Xue J., Bartlett J.D.;

RT "Expression and localization of membrane type 1 matrix

RT metalloproteinase in tooth tissues.";

RL Matrix Biol. 17:501-511(1998).

CC -!- FUNCTION: Seems to specifically activate progelatinase A. May thus  
CC trigger invasion by tumor cells by activating progelatinase A on  
CC the tumor cell surface (By similarity). May play a role in the  
CC biomineralization of enamel and dentin.

CC -!- CATALYTIC ACTIVITY: Endopeptidase activity. Activates  
CC progelatinase A by cleavage of the propeptide at 37-Asn-|-Leu-38.  
CC Other bonds hydrolyzed include 35-Gly-|-Ile-36 in the propeptide  
CC of collagenase 3, and 341-Asn-|-Phe-342, 441-Asp-|-Leu-442 and

CC 354-Gln-|-Thr-355 in the aggrecan interglobular domain.  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit, calcium (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).  
 CC -!- TISSUE SPECIFICITY: Highly expressed in developing tooth tissues.  
 CC -!- SIMILARITY: Belongs to peptidase family M10A.  
 CC -!- SIMILARITY: Contains 1 hemopexin-like domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF067419; AAD38324.1; -.  
 DR HSSP; P08254; 1HFS.  
 DR MEROPS; M10.014; -.  
 DR InterPro; IPR000585; Hemopexin.  
 DR InterPro; IPR001818; Pept\_M10A\_M12B.  
 DR InterPro; IPR006025; Pept\_M\_Zn\_BS.  
 DR InterPro; IPR006026; Peptidase\_M.  
 DR Pfam; PF00045; hemopexin; 4.  
 DR Pfam; PF00413; Peptidase\_M10; 1.  
 DR Pfam; PF03933; Peptidase\_M10\_N; 1.  
 DR PRINTS; PR00138; MATRIXIN.  
 DR SMART; SM00120; HX; 4.  
 DR SMART; SM00235; ZnMc; 1.  
 DR PROSITE; PS00024; HEMOPEXIN; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 DR PROSITE; PS00546; CYSTEINE\_SWITCH; 1.  
 KW Hydrolase; Metalloprotease; Zinc; Zymogen; Calcium; Signal;  
 KW Transmembrane.  
 FT SIGNAL 1 28 POTENTIAL.  
 FT PROPEP 29 109 ACTIVATION PEPTIDE.  
 FT CHAIN 110 580 MATRIX METALLOPROTEINASE-14.  
 FT DOMAIN 110 539 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 540 560 POTENTIAL.  
 FT DOMAIN 561 580 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 314 509 HEMOPEXIN-LIKE.  
 FT SITE 91 91 CYSTEINE SWITCH (POTENTIAL).  
 FT METAL 237 237 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 238 238 BY SIMILARITY.  
 FT METAL 241 241 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 247 247 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DISULFID 317 506 BY SIMILARITY.  
 SQ SEQUENCE 580 AA; 65934 MW; B7B2C2C569A96CAC CRC64;

Query Match 75.6%; Score 31; DB 1; Length 580;  
 Best Local Similarity 62.5%; Pred. No. 54;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 :::||| |  
 Db 176 MIFFAEGF 183

RESULT 20

MM14\_HUMAN

ID MM14\_HUMAN STANDARD; PRT; 582 AA.  
AC P50281; Q92678;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Matrix metalloproteinase-14 precursor (EC 3.4.24.80) (MMP-14)  
DE (Membrane-type matrix metalloproteinase 1) (MT-MMP 1) (MTMMP1)  
DE (Membrane-type-1 matrix metalloproteinase) (MT1-MMP) (MT1MMP) (MMP-  
DE X1).  
GN MMP14.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=94286011; PubMed=8015608;  
RA Sato H., Takino T., Okada Y., Cao J., Shinagawa A., Yamamoto E.,  
RA Seiki M.;  
RT "A matrix metalloproteinase expressed on the surface of invasive  
RT tumour cells.";  
RL Nature 370:61-65(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=95237627; PubMed=7721107;  
RA Takino T., Sato H., Yamamoto E., Seiki M.;  
RT "Cloning of a human gene potentially encoding a novel matrix  
RT metalloproteinase having a C-terminal transmembrane domain.";  
RL Gene 155:293-298(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95224014; PubMed=7708715;  
RA Okada A., Bellocq J.P., Rouyer N., Chenard M.P., Rio M.C.,  
RA Chambon P., Basset P.;  
RT "Membrane-type matrix metalloproteinase (MT-MMP) gene is expressed in  
RT stromal cells of human colon, breast, and head and neck carcinomas.";  
RL Proc. Natl. Acad. Sci. U.S.A. 92:2730-2734(1995).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=95377289; PubMed=7649159;  
RA Will H., Hinzmann B.;  
RT "cDNA sequence and mRNA tissue distribution of a novel human matrix  
RT metalloproteinase with a potential transmembrane segment.";  
RL Eur. J. Biochem. 231:602-608(1995).  
RN [5]  
RP SEQUENCE FROM N.A.  
RA Luo G.X., Reisfeld R.A., Strongin A.Y.;  
RL Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RA Lohi J.L., Westermarck J., Kaehaeri V.M., Keski-Oja J.;  
RL Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.



```

CC  -!- FUNCTION: Seems to specifically activate progelatinase A. May thus
CC      trigger invasion by tumor cells by activating progelatinase A on
CC      the tumor cell surface.
CC  -!- CATALYTIC ACTIVITY: Endopeptidase activity. Activates
CC      progelatinase A by cleavage of the propeptide at 37-Asn-|-Leu-38.
CC      Other bonds hydrolyzed include 35-Gly-|-Ile-36 in the propeptide
CC      of collagenase 3, and 341-Asn-|-Phe-342, 441-Asp-|-Leu-442 and
CC      354-Gln-|-Thr-355 in the aggrecan interglobular domain.
CC  -!- COFACTOR: Binds 1 zinc ion per subunit, calcium (By similarity).
CC  -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC  -!- TISSUE SPECIFICITY: In stromal cells of colon, breast, and head
CC      and neck.
CC  -!- SIMILARITY: Belongs to peptidase family M10A.
CC  -!- SIMILARITY: Contains 1 hemopexin-like domain.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; D26512; BAA05519.1; -.
DR  EMBL; X83535; CAA58519.1; -.
DR  EMBL; Z48481; CAA88372.1; -.
DR  EMBL; U41078; AAA83770.1; -.
DR  EMBL; X90925; CAA62432.1; -.
DR  PIR; I38028; I38028.
DR  PDB; 1BQQ; 23-AUG-01.
DR  PDB; 1BUV; 02-SEP-99.
DR  MEROPS; M10.014; -.
DR  Genew; HGNC:7160; MMP14.
DR  MIM; 600754; -.
DR  GO; GO:0005887; C:integral to plasma membrane; TAS.
DR  GO; GO:0004222; F:metalloendopeptidase activity; TAS.
DR  GO; GO:0008270; F:zinc ion binding; TAS.
DR  GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
DR  InterPro; IPR000585; Hemopexin.
DR  InterPro; IPR001818; Pept_M10A_M12B.
DR  InterPro; IPR006025; Pept_M_Zn_BS.
DR  InterPro; IPR006026; Peptidase_M.
DR  Pfam; PF00045; hemopexin; 4.
DR  Pfam; PF00413; Peptidase_M10; 1.
DR  Pfam; PF03933; Peptidase_M10_N; 1.
DR  PRINTS; PR00138; MATRIXIN.
DR  SMART; SM00120; HX; 4.
DR  SMART; SM00235; ZnMc; 1.
DR  PROSITE; PS00024; HEMOPEXIN; 1.
DR  PROSITE; PS00142; ZINC_PROTEASE; 1.
DR  PROSITE; PS00546; CYSTEINE_SWITCH; 1.
KW  Hydrolase; Metalloprotease; Zinc; Zymogen; Calcium; Signal;
KW  Transmembrane; 3D-structure.
FT  SIGNAL      1      20      POTENTIAL.
FT  PROPEP      21     111     ACTIVATION PEPTIDE.
FT  CHAIN       112     582     MATRIX METALLOPROTEINASE-14.
FT  DOMAIN      112     541     EXTRACELLULAR (POTENTIAL).

```

FT	TRANSMEM	542	562	POTENTIAL.
FT	DOMAIN	563	582	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	316	511	HEMOPEXIN-LIKE.
FT	SITE	93	93	CYSTEINE SWITCH (POTENTIAL).
FT	METAL	239	239	ZINC (CATALYTIC) (BY SIMILARITY).
FT	ACT_SITE	240	240	BY SIMILARITY.
FT	METAL	243	243	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL	249	249	ZINC (CATALYTIC) (BY SIMILARITY).
FT	DISULFID	319	508	BY SIMILARITY.
FT	CONFLICT	338	338	K -> E (IN REF. 2, 4, 5 AND 6).
FT	CONFLICT	500	500	S -> P (IN REF. 6).
SQ	SEQUENCE	582 AA; 65883 MW; EFCEDCE6A41116F9 CRC64;		

Query Match 75.6%; Score 31; DB 1; Length 582;  
 Best Local Similarity 62.5%; Pred. No. 55;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LVFFAEDF 8  
 ::|||||  
 Db 178 MIFFAEGF 185

# RESULT 21

## MM14\_RABIT

ID MM14\_RABIT STANDARD; PRT; 582 AA.  
 AC Q95220; P79225;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Matrix metalloproteinase-14 precursor (EC 3.4.24.80) (MMP-14)  
 DE (Membrane-type matrix metalloproteinase 1) (MT-MMP 1) (MTMMP1)  
 DE (Membrane-type-1 matrix metalloproteinase) (MT1-MMP) (MT1MMP).  
 GN MMP14.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=New Zealand white; TISSUE=Vascular smooth muscle;  
 RA Wang H., Keiser J.;  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE OF 1-572 FROM N.A.  
 RC STRAIN=New Zealand white;  
 RA Sato T.;  
 RL Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: Seems to specifically activate progelatinase A. May thus  
 CC trigger invasion by tumor cells by activating progelatinase A on  
 CC the tumor cell surface (By similarity).  
 CC -!- CATALYTIC ACTIVITY: Endopeptidase activity. Activates  
 CC progelatinase A by cleavage of the propeptide at 37-Asn-|-Leu-38.  
 CC Other bonds hydrolyzed include 35-Gly-|-Ile-36 in the propeptide  
 CC of collagenase 3, and 341-Asn-|-Phe-342, 441-Asp-|-Leu-442 and  
 CC 354-Gln-|-Thr-355 in the aggrecan interglobular domain.  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit, calcium (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).

```

CC  -!- SIMILARITY: Belongs to peptidase family M10A.
CC  -!- SIMILARITY: Contains 1 hemopexin-like domain.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; U83918; AAB41500.1; -.
DR  EMBL; U73940; AAD13803.1; -.
DR  HSSP; P08254; 1HFS.
DR  MEROPS; M10.014; -.
DR  InterPro; IPR000585; Hemopexin.
DR  InterPro; IPR001818; Pept_M10A_M12B.
DR  InterPro; IPR006025; Pept_M_Zn_BS.
DR  InterPro; IPR006026; Peptidase_M.
DR  Pfam; PF00045; hemopexin; 4.
DR  Pfam; PF00413; Peptidase_M10; 1.
DR  Pfam; PF03933; Peptidase_M10_N; 1.
DR  PRINTS; PR00138; MATRIXIN.
DR  SMART; SM00120; HX; 4.
DR  SMART; SM00235; ZnMc; 1.
DR  PROSITE; PS00024; HEMOPEXIN; 1.
DR  PROSITE; PS00142; ZINC_PROTEASE; 1.
DR  PROSITE; PS00546; CYSTEINE_SWITCH; 1.
KW  Hydrolase; Metalloprotease; Zinc; Zymogen; Calcium; Signal;
KW  Transmembrane.
FT  SIGNAL          1      20      POTENTIAL.
FT  PROPEP          21     111     ACTIVATION PEPTIDE.
FT  CHAIN           112     582     MATRIX METALLOPROTEINASE-14.
FT  DOMAIN          112     541     EXTRACELLULAR (POTENTIAL).
FT  TRANSMEM        542     562     POTENTIAL.
FT  DOMAIN          563     582     CYTOPLASMIC (POTENTIAL).
FT  DOMAIN          316     511     HEMOPEXIN-LIKE.
FT  SITE            93      93      CYSTEINE SWITCH (POTENTIAL).
FT  METAL           239     239     ZINC (CATALYTIC) (BY SIMILARITY).
FT  ACT_SITE        240     240     BY SIMILARITY.
FT  METAL           243     243     ZINC (CATALYTIC) (BY SIMILARITY).
FT  METAL           249     249     ZINC (CATALYTIC) (BY SIMILARITY).
FT  DISULFID        319     508     BY SIMILARITY.
FT  CONFLICT        29      29      Q -> K (IN REF. 2).
FT  CONFLICT       268     268     K -> N (IN REF. 2).
FT  CONFLICT       270     270     L -> V (IN REF. 2).
FT  CONFLICT       275     275     E -> D (IN REF. 2).
FT  CONFLICT       292     296     RCLLN -> KMPPP (IN REF. 2).
FT  CONFLICT       298     300     GQP -> RTT (IN REF. 2).
FT  CONFLICT       302     308     GLLFRIS -> RTFIPDK (IN REF. 2).
FT  CONFLICT       310     310     G -> R (IN REF. 2).
FT  CONFLICT       317     317     K -> N (IN REF. 2).
FT  CONFLICT       329     329     F -> L (IN REF. 2).
FT  CONFLICT       360     360     L -> F (IN REF. 2).
SQ  SEQUENCE        582 AA;  65963 MW;  844624B0AF1B6812 CRC64;

```

Query Match

75.6%; Score 31; DB 1; Length 582;

Best Local Similarity 62.5%; Pred. No. 55;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
::|||||  
Db 178 MIFFAEGF 185

RESULT 22

YRT1\_CAEEL

ID YRT1\_CAEEL STANDARD; PRT; 622 AA.  
AC Q10044;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Hypothetical 71.6 kDa protein T07A5.1 in chromosome III.  
GN T07A5.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Buck D.;  
RL Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.  
CC -!- SIMILARITY: IN THE C-TERMINAL, TO C.ELEGANS W02B3.4 AND W02B3.5.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; Z48055; CAA88133.1; -.  
DR PIR; T24632; T24632.  
DR WormPep; T07A5.1; CE01646.  
DR InterPro; IPR007074; LicD.  
DR Pfam; PF04991; LicD; 1.  
KW Hypothetical protein.  
FT DOMAIN 5 8 POLY-LEU.  
FT DOMAIN 148 154 POLY-SER.  
SQ SEQUENCE 622 AA; 71623 MW; D3E4BDAB67DC3AD0 CRC64;

Query Match 75.6%; Score 31; DB 1; Length 622;  
Best Local Similarity 75.0%; Pred. No. 58;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
| |||||  
Db 514 LAMFAEDF 521

RESULT 23

MTN2\_HUMAN

ID MTN2\_HUMAN STANDARD; PRT; 956 AA.  
 AC O00339; Q9NSZ1;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Matrilin-2 precursor.  
 GN MATN2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).  
 RX PubMed=11124542;  
 RA Muratoglu S., Krysan K., Balazs M., Sheng H., Zakany R., Modis L.,  
 RA Kiss I., Deak F.;  
 RT "Primary structure of human matrilin-2, chromosome location of the  
 RT MATN2 gene and conservation of an AT-AC intron in matrilin genes."  
 RL Cytogenet. Cell Genet. 90:323-327(2000).  
 RN [2]  
 RP SEQUENCE OF 644-956 FROM N.A.  
 RX MEDLINE=97238863; PubMed=9083061;  
 RA Deak F., Piecha D., Bachrati C., Paulsson M., Kiss I.;  
 RT "Primary structure and expression of matrilin-2, the closest relative  
 RT of cartilage matrix protein within the von Willebrand factor type A-  
 RT like module superfamily."  
 RL J. Biol. Chem. 272:9268-9274(1997).  
 RN [3]  
 RP SEQUENCE OF 244-956 FROM N.A.  
 RC TISSUE=Testis;  
 RA Duesterhoeft A., Lauber J., Mewes H.-W., Gassenhuber J., Wiemann S.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: Involved in matrix assembly (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=Long;  
 CC IsoId=O00339-1; Sequence=Displayed;  
 CC Name=Short;  
 CC IsoId=O00339-2; Sequence=VSP\_001399;  
 CC -!- SIMILARITY: Contains 10 EGF-like domains.  
 CC -!- SIMILARITY: Contains 2 VWFA domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U69263; AAC51260.2; -.  
 DR EMBL; AL137638; CAB70853.1; ALT\_INIT.  
 DR HSSP; P01132; 1EGF.  
 DR Genew; HGNC:6908; MATN2.  
 DR MIM; 602108; -.  
 DR GO; GO:0005578; C:extracellular matrix; NAS.

DR InterPro; IPR000152; Asx\_hydroxyl\_S.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF00008; EGF; 10.  
 DR Pfam; PF00092; vwa; 2.  
 DR PRINTS; PR00453; VWFADOMAIN.  
 DR SMART; SM00179; EGF\_CA; 1.  
 DR SMART; SM00327; VWA; 2.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 9.  
 DR PROSITE; PS01186; EGF\_2; 9.  
 DR PROSITE; PS50026; EGF\_3; 4.  
 DR PROSITE; PS50234; VWFA; 2.  
 KW EGF-like domain; Signal; Glycoprotein; Repeat; Coiled coil;  
 KW Alternative splicing.  
 FT SIGNAL 1 23 POTENTIAL.  
 FT CHAIN 24 956 MATRILIN-2.  
 FT DOMAIN 57 232 VWFA 1.  
 FT DOMAIN 238 278 EGF-LIKE 1.  
 FT DOMAIN 279 319 EGF-LIKE 2.  
 FT DOMAIN 320 360 EGF-LIKE 3.  
 FT DOMAIN 361 401 EGF-LIKE 4.  
 FT DOMAIN 402 442 EGF-LIKE 5.  
 FT DOMAIN 443 483 EGF-LIKE 6.  
 FT DOMAIN 484 524 EGF-LIKE 7.  
 FT DOMAIN 525 565 EGF-LIKE 8.  
 FT DOMAIN 566 606 EGF-LIKE 9.  
 FT DOMAIN 607 647 EGF-LIKE 10.  
 FT DOMAIN 655 830 VWFA 2.  
 FT DOMAIN 917 955 COILED COIL (POTENTIAL).  
 FT DISULFID 242 253 BY SIMILARITY.  
 FT DISULFID 249 262 BY SIMILARITY.  
 FT DISULFID 264 277 BY SIMILARITY.  
 FT DISULFID 283 294 BY SIMILARITY.  
 FT DISULFID 290 303 BY SIMILARITY.  
 FT DISULFID 305 318 BY SIMILARITY.  
 FT DISULFID 324 335 BY SIMILARITY.  
 FT DISULFID 331 344 BY SIMILARITY.  
 FT DISULFID 346 359 BY SIMILARITY.  
 FT DISULFID 365 376 BY SIMILARITY.  
 FT DISULFID 372 385 BY SIMILARITY.  
 FT DISULFID 387 400 BY SIMILARITY.  
 FT DISULFID 406 417 BY SIMILARITY.  
 FT DISULFID 413 426 BY SIMILARITY.  
 FT DISULFID 428 441 BY SIMILARITY.  
 FT DISULFID 447 458 BY SIMILARITY.  
 FT DISULFID 454 467 BY SIMILARITY.  
 FT DISULFID 469 482 BY SIMILARITY.  
 FT DISULFID 488 499 BY SIMILARITY.  
 FT DISULFID 495 508 BY SIMILARITY.  
 FT DISULFID 510 523 BY SIMILARITY.  
 FT DISULFID 529 540 BY SIMILARITY.  
 FT DISULFID 536 549 BY SIMILARITY.  
 FT DISULFID 551 564 BY SIMILARITY.  
 FT DISULFID 570 581 BY SIMILARITY.  
 FT DISULFID 577 590 BY SIMILARITY.  
 FT DISULFID 592 605 BY SIMILARITY.

FT DISULFID 611 622 BY SIMILARITY.  
 FT DISULFID 618 631 BY SIMILARITY.  
 FT DISULFID 633 646 BY SIMILARITY.  
 FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 890 890 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 861 879 Missing (in isoform Short).  
 FT /FTId=VSP\_001399.  
 FT CONFLICT 356 356 K -> E (IN REF. 3).  
 FT CONFLICT 594 594 V -> E (IN REF. 3).  
 FT CONFLICT 644 644 R -> G (IN REF. 2).  
 FT CONFLICT 755 755 F -> L (IN REF. 3).  
 FT CONFLICT 935 935 L -> F (IN REF. 2).  
 SQ SEQUENCE 956 AA; 106840 MW; 826B7F347178FC80 CRC64;

Query Match 75.6%; Score 31; DB 1; Length 956;  
 Best Local Similarity 71.4%; Pred. No. 90;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
 :|:||||  
 Db 814 LFYAEDF 820

#### RESULT 24

##### MTN2\_MOUSE

ID MTN2\_MOUSE STANDARD; PRT; 956 AA.  
 AC O08746;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Matrilin-2 precursor.  
 GN MATN2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Limb;  
 RX MEDLINE=97238863; PubMed=9083061;  
 RA Deak F., Piecha D., Bachrati C., Paulsson M., Kiss I.;  
 RT "Primary structure and expression of matrilin-2, the closest relative  
 RT of cartilage matrix protein within the von Willebrand factor type A-  
 RT like module superfamily."  
 RL J. Biol. Chem. 272:9268-9274(1997).  
 CC -!- FUNCTION: Involved in matrix assembly (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Detected in a variety of organs, including  
 CC calvaria, uterus, heart and brain, as well as fibroblast and  
 CC osteoblast cell lines.  
 CC -!- SIMILARITY: Contains 10 EGF-like domains.  
 CC -!- SIMILARITY: Contains 2 VWFA domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; U69262; AAC53163.1; -.  
 DR HSSP; P05099; 1AQ5.  
 DR MGD; MGI:109613; Matn2.  
 DR GO; GO:0005578; C:extracellular matrix; IDA.  
 DR InterPro; IPR000152; Asx\_hydroxyl\_S.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF00008; EGF; 10.  
 DR Pfam; PF00092; vwa; 2.  
 DR PRINTS; PR00453; VWFADOMAIN.  
 DR SMART; SM00181; EGF; 10.  
 DR SMART; SM00327; VWA; 2.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 7.  
 DR PROSITE; PS01186; EGF\_2; 9.  
 DR PROSITE; PS50234; VWFA; 2.  
 KW EGF-like domain; Signal; Glycoprotein; Repeat; Coiled coil.

FT	SIGNAL	1	23	POTENTIAL.
FT	CHAIN	24	956	MATRILIN-2.
FT	DOMAIN	57	232	VWFA 1.
FT	DOMAIN	238	278	EGF-LIKE 1.
FT	DOMAIN	279	319	EGF-LIKE 2.
FT	DOMAIN	320	360	EGF-LIKE 3.
FT	DOMAIN	361	401	EGF-LIKE 4.
FT	DOMAIN	402	442	EGF-LIKE 5.
FT	DOMAIN	443	483	EGF-LIKE 6.
FT	DOMAIN	484	524	EGF-LIKE 7.
FT	DOMAIN	525	565	EGF-LIKE 8.
FT	DOMAIN	566	606	EGF-LIKE 9.
FT	DOMAIN	607	647	EGF-LIKE 10.
FT	DOMAIN	655	830	VWFA 2.
FT	DOMAIN	917	955	COILED COIL (POTENTIAL).
FT	DISULFID	242	253	BY SIMILARITY.
FT	DISULFID	249	262	BY SIMILARITY.
FT	DISULFID	264	277	BY SIMILARITY.
FT	DISULFID	283	294	BY SIMILARITY.
FT	DISULFID	290	303	BY SIMILARITY.
FT	DISULFID	305	318	BY SIMILARITY.
FT	DISULFID	324	335	BY SIMILARITY.
FT	DISULFID	331	344	BY SIMILARITY.
FT	DISULFID	346	359	BY SIMILARITY.
FT	DISULFID	365	376	BY SIMILARITY.
FT	DISULFID	372	385	BY SIMILARITY.
FT	DISULFID	387	400	BY SIMILARITY.
FT	DISULFID	406	417	BY SIMILARITY.
FT	DISULFID	413	426	BY SIMILARITY.
FT	DISULFID	428	441	BY SIMILARITY.
FT	DISULFID	447	458	BY SIMILARITY.
FT	DISULFID	454	467	BY SIMILARITY.
FT	DISULFID	469	482	BY SIMILARITY.
FT	DISULFID	488	499	BY SIMILARITY.
FT	DISULFID	495	508	BY SIMILARITY.



FT	DISULFID	510	523	BY SIMILARITY.
FT	DISULFID	529	540	BY SIMILARITY.
FT	DISULFID	536	549	BY SIMILARITY.
FT	DISULFID	551	564	BY SIMILARITY.
FT	DISULFID	570	581	BY SIMILARITY.
FT	DISULFID	577	590	BY SIMILARITY.
FT	DISULFID	592	605	BY SIMILARITY.
FT	DISULFID	611	622	BY SIMILARITY.
FT	DISULFID	618	631	BY SIMILARITY.
FT	DISULFID	633	646	BY SIMILARITY.
FT	CARBOHYD	221	221	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	890	890	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	956 AA; 106779 MW; 3E4C22770B6E6EEE CRC64;		

Query Match 75.6%; Score 31; DB 1; Length 956;  
 Best Local Similarity 71.4%; Pred. No. 90;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 VFFAEDF 8  
 :|:||||  
 Db 814 LFYAEDF 820

RESULT 25

FAB1\_SCHPO

ID FAB1\_SCHPO STANDARD; PRT; 1932 AA.  
 AC O59722; O43072;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Phosphatidylinositol-4-phosphate 5-kinase fab1 (EC 2.7.1.68) (1-phosphatidylinositol-4-phosphate kinase) (PIP5K) (PtdIns(4)P-5-kinase)  
 DE (Diphosphoinositide kinase).  
 GN FAB1 OR STE12 OR SPBC3E7.01 OR SPBC6B1.11C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.  
 RX MEDLINE=21893055; PubMed=11895483;  
 RA Morishita M., Morimoto F., Kitamura K., Koga T., Fukui Y., Maekawa H.,  
 RA Yamashita I., Shimoda C.;  
 RT "Phosphatidylinositol 3-phosphate 5-kinase is required for the  
 RT cellular response to nutritional starvation and mating pheromone  
 RT signals in Schizosaccharomyces pombe."  
 RL Genes Cells 7:199-215(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,

RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbino-witsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of *Schizosaccharomyces pombe*."  
 RL Nature 415:871-880(2002).  
 CC -!- FUNCTION: Catalyzes the phosphorylation of phosphatidylinositol-4-  
 CC phosphate on the fifth hydroxyl of the myo-inositol ring, to form  
 CC phosphatidylinositol-4,5-bisphosphate. Required for survival under  
 CC conditions of nitrogen starvation. May have a role in the  
 CC secretion of pheromone peptides.  
 CC -!- CATALYTIC ACTIVITY: ATP + 1-phosphatidyl-1D-myo-inositol 4-  
 CC monophosphate = ADP + 1-phosphatidyl-1D-myo-inositol 4,5-  
 CC bisphosphate.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic. During cell fusion of opposite  
 CC mating types, migrates to the isthmus.  
 CC -!- SIMILARITY: BELONGS TO THE PTDINS(4)P-5-KINASE FAMILY.  
 CC -!- SIMILARITY: Contains 1 FYVE-type zinc finger.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AL023534; CAA19005.1; -.  
 DR EMBL; AL021838; CAA17054.1; -.  
 DR PIR; T40375; T40375.  
 DR GeneDB\_SPombe; SPBC3E7.01; -.  
 DR GO; GO:0006629; P:lipid metabolism; ISS.  
 DR GO; GO:0007165; P:signal transduction; ISS.  
 DR InterPro; IPR002498; PIP5K.  
 DR InterPro; IPR000306; Znf\_FYVE.  
 DR Pfam; PF01363; FYVE; 1.  
 DR SMART; SM00064; FYVE; 1.  
 DR SMART; SM00330; PIPKc; 1.  
 DR PROSITE; PS50178; ZF\_FYVE; 1.  
 KW Transferase; Kinase; Zinc-finger; Zinc.  
 FT ZN\_FING 60 119 FYVE-TYPE.  
 SQ SEQUENCE 1932 AA; 220132 MW; B9826516D4504229 CRC64;

Query Match 75.6%; Score 31; DB 1; Length 1932;  
Best Local Similarity 85.7%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
|||||  
Db 1662 VFFAEQF 1668

RESULT 26

MOR2\_SCHPO

ID MOR2\_SCHPO STANDARD; PRT; 2196 AA.  
AC Q9HDV6;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Cell polarity protein mor2 (Morphological round protein 2).  
GN MOR2 OR SPBP19A11.04C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomyces.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.  
RC STRAIN=972;  
RX MEDLINE=22220058; PubMed=12234926;  
RA Hirata D., Kishimoto N., Suda M., Sogabe Y., Nakagawa S., Yoshida Y.,  
RA Sakai K., Mizunuma M., Miyakawa T., Ishiguro J., Toda T.;  
RT "Fission yeast Mor2/Cps12, a protein similar to Drosophila Furry, is  
RT essential for cell morphogenesis and its mutation induces  
RT Wee1-dependent G2 delay.";  
RL EMBO J. 21:4863-4874(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,

RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of *Schizosaccharomyces pombe*."  
 RL Nature 415:871-880(2002).  
 CC -!- FUNCTION: Required for the maintenance of cell polarity. Has a  
 CC role in localizing F-actin at the cell tips.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic and membrane-associated; found  
 CC at the cell tips and at the site of septum formation.  
 CC -!- SIMILARITY: TO YEAST TAO3/PAG1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB093007; BAC20935.1; -.  
 DR EMBL; AL512495; CAC19754.1; -.  
 DR GeneDB\_SPombe; SPBP19A11.04c; -.  
 SQ SEQUENCE 2196 AA; 251195 MW; AAF2698DB00DE4CF CRC64;

Query Match 75.6%; Score 31; DB 1; Length 2196;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VFFAED 7  
 |||||  
 Db 2014 VFFAED 2019

# RESULT 27

Y681\_CHLPN

ID Y681\_CHLPN STANDARD; PRT; 224 AA.  
 AC Q9Z7M3; Q9JQF7;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Protein CPn0681/CP0066/CPj0681/CpB0708.  
 GN CPN0681 OR CP0066 OR CPJ0681 OR CPB0708.  
 OS *Chlamydia pneumoniae* (*Chlamydophila pneumoniae*).  
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CWL029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*."  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RC STRAIN=AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
 RT pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CWL029 from USA.";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TW-183;  
 RA Geng M.M., Schuhmacher A., Muehldorfer I., Bensch K.W., Schaefer K.P.,  
 RA Schneider S., Pohl T., Essig A., Marre R., Melchers K.;  
 RT "The genome sequence of Chlamydia pneumoniae TW183 and comparison with  
 RT other Chlamydia strains based on whole genome sequence analysis.";  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: Belongs to the UPF0111 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE001650; AAD18820.1; -.  
 DR EMBL; AE002170; AAF37955.1; -.  
 DR EMBL; AP002547; BAA98888.1; -.  
 DR EMBL; AE017159; AAP98637.1; -.  
 DR PIR; E72049; E72049.  
 DR PIR; F86575; F86575.  
 DR PHCI-2DPAGE; Q927M3; -.  
 DR TIGR; CP0066; -.  
 DR InterPro; IPR002727; DUF47.  
 DR Pfam; PF01865; PhoU\_div; 1.  
 DR TIGRFAMs; TIGR00153; TIGR00153; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 224 AA; 25887 MW; 82D1AF044CD11EA0 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 224;  
 Best Local Similarity 57.1%; Pred. No. 35;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
 :||: ||

## RESULT 28

## SYE\_WIGBR

ID SYE\_WIGBR STANDARD; PRT; 473 AA.  
 AC Q8D375;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Glutamyl-tRNA synthetase (EC 6.1.1.17) (Glutamate--tRNA ligase)  
 DE (GluRS).  
 GN GLTX OR WIGBR1260.  
 OS Wigglesworthia glossinidia brevipalpis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Wigglesworthia.  
 OX NCBI\_TaxID=36870;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22297718; PubMed=12219091;  
 RA Akman L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,  
 RA Aksoy S.;  
 RT "Genome sequence of the endocellular obligate symbiont of tsetse  
 RT flies, Wigglesworthia glossinidia.";  
 RL Nat. Genet. 32:402-407(2002).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +  
 CC diphosphate + L-glutamyl-tRNA(Glu).  
 CC -!- SUBUNIT: Monomer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB063521; BAC24272.1; -.  
 DR HAMAP; MF\_00022; -; 1.  
 DR InterPro; IPR004527; GltX\_bact.  
 DR InterPro; IPR000924; Glu\_tRNA-synt\_1c.  
 DR InterPro; IPR008925; tRNA-synt\_bind.  
 DR InterPro; IPR001412; tRNA-synt\_I.  
 DR Pfam; PF00749; tRNA-synt\_1c; 1.  
 DR PRINTS; PR00987; TRNASYNTHGLU.  
 DR TIGRFAMs; TIGR00464; gltX\_bact; 1.  
 DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; FALSE\_NEG.  
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 KW Complete proteome.  
 FT SITE 9 19 "HIGH" REGION.  
 FT SITE 237 241 "KMSKS" REGION.  
 FT BINDING 240 240 ATP (BY SIMILARITY).  
 SQ SEQUENCE 473 AA; 55832 MW; E20D10F9F7CFA7C9 CRC64;

Best Local Similarity 62.5%; Pred. No. 73;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
::|| |||  
Db 361 ILFFYEDF 368

RESULT 29

YQP4\_CAEEEL

ID YQP4\_CAEEEL STANDARD; PRT; 529 AA.  
AC Q09531;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical protein F07F6.4 in chromosome III.  
GN F07F6.4.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Chissoe S.;  
RL Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP REVISIONS.  
RA Waterston R.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.  
CC -!- FUNCTION: GTPase-activating protein for the ADP ribosylation  
CC factor family (Potential).  
CC -!- SIMILARITY: Contains 1 Arf-GAP domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U23486; AAL38955.1; -.  
DR WormPep; F07F6.4; CE29955.  
DR InterPro; IPR001164; hRIP\_like.  
DR Pfam; PF01412; ArfGap; 1.  
DR PRINTS; PR00405; REVINTRACTNG.  
DR SMART; SM00105; ArfGap; 1.  
DR PROSITE; PS50115; ARFGAP; 1.  
KW Hypothetical protein; GTPase activation; Zinc-finger.  
FT DOMAIN 13 129 ARF-GAP.  
FT ZN\_FING 28 51 C4-TYPE.  
FT DOMAIN 375 378 POLY-SER.  
FT DOMAIN 404 411 POLY-SER.  
SQ SEQUENCE 529 AA; 57911 MW; FE67BC185F25ECC3 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 529;

Best Local Similarity 83.3%; Pred. No. 82;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 FFAEDF 8  
|||:|  
Db 153 FFAQDF 158

RESULT 30

GRAU\_DROME

ID GRAU\_DROME STANDARD; PRT; 570 AA.  
AC Q9U405; Q9W2N7;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Transcription factor grauzone.  
GN GRAU OR CG3282.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND MUTANTS QQ36 AND  
RP QE70.  
RX MEDLINE=20148536; PubMed=10683177;  
RA Chen B., Harms E., Chu T., Henrion G., Strickland S.;  
RT "Completion of meiosis in Drosophila oocytes requires transcriptional  
RT control by grauzone, a new zinc finger protein.";  
RL Development 127:1243-1251(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,



RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [3]  
 RP REVISIONS.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 RT systematic review.";  
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley; TISSUE=Head;  
 RX MEDLINE=22426066; PubMed=12537569;  
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
 RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,  
 RA Rubin G.M., Celniker S.E.;  
 RT "A *Drosophila* full-length cDNA resource.";  
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
 RN [5]  
 RP FUNCTION, AND DNA-BINDING.  
 RX MEDLINE=20384592; PubMed=10924478;  
 RA Harms E., Chu T., Henrion G., Strickland S.;  
 RT "The only function of Grauzone required for *Drosophila* oocyte meiosis  
 RT is transcriptional activation of the cortex gene.";  
 RL Genetics 155:1831-1839(2000).  
 CC -!- FUNCTION: Transcription factor involved in female meiosis.  
 CC Specifically required to bind to the promoter region of the Cortex  
 CC gene and activate its expression in ovaries.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- TISSUE SPECIFICITY: Present throughout embryogenesis, in larvae,  
 CC in ovaries and other tissues of adult females, and in adult males.  
 CC In ovaries, it is expressed in both follicle cells and nurse  
 CC cells. Expressed from stage 9 in the germarium. Weakly or not  
 CC expressed in the oocyte nucleus.  
 CC -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.  
 CC -!- SIMILARITY: Contains 7 C2H2-type zinc fingers.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AF208016; AAF25356.1; -.  
 DR EMBL; AE003452; AAF46653.2; -.  
 DR EMBL; AY047537; AAK77269.1; -.  
 DR GermOnline; 209995; -.  
 DR FlyBase; FBgn0001133; grau.  
 DR GO; GO:0005634; C:nucleus; IDA.  
 DR GO; GO:0003704; F:specific RNA polymerase II transcription fa. . .; IMP.  
 DR GO; GO:0007143; P:female meiosis; IMP.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF00096; zf-C2H2; 8.  
 DR SMART; SM00355; Znf\_C2H2; 8.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 7.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 7.  
 KW Meiosis; DNA-binding; Transcription regulation; Activator;  
 KW Metal-binding; Zinc; Zinc-finger; Repeat.  
 FT ZN\_FING 296 319 C2H2-TYPE 1.  
 FT ZN\_FING 327 349 C2H2-TYPE 2.  
 FT ZN\_FING 387 409 C2H2-TYPE 3.  
 FT ZN\_FING 417 441 C2H2-TYPE 4.  
 FT ZN\_FING 448 471 C2H2-TYPE 5.  
 FT ZN\_FING 476 498 C2H2-TYPE 6.  
 FT ZN\_FING 504 527 C2H2-TYPE 7.  
 FT DOMAIN 293 570 DNA-BINDING.  
 FT DOMAIN 383 399 BIPARTITE NUCLEAR LOCALIZATION SIGNAL 1  
 FT (POTENTIAL).  
 FT DOMAIN 523 539 BIPARTITE NUCLEAR LOCALIZATION SIGNAL 2  
 FT (POTENTIAL).  
 FT DOMAIN 146 172 ASP/GLU-RICH (ACIDIC).  
 FT MUTAGEN 298 298 C->Y: IN QQ36; DEFECTS IN FEMALE MEIOSIS.  
 FT ABOLISHES DNA-BINDING ACTIVITY.  
 FT MUTAGEN 493 493 E->K: IN QE70; DEFECTS IN FEMALE MEIOSIS.  
 FT ABOLISHES DNA-BINDING ACTIVITY.  
 SQ SEQUENCE 570 AA; 65967 MW; 33B550B964C1EA7C CRC64;

Query Match 73.2%; Score 30; DB 1; Length 570;  
 Best Local Similarity 71.4%; Pred. No. 88;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
 :| ||||  
 Db 246 IFLAEDF 252

RESULT 31  
 LICR\_BACSU  
 ID LICR\_BACSU STANDARD; PRT; 641 AA.  
 AC P46321;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Probable *licABCH* operon regulator.  
 GN LICR OR CELR OR BSU38600.  
 OS *Bacillus subtilis*.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; *Bacillus*.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=97144536; PubMed=8990303;  
 RA Tobisch S., Glaser P., Krueger S., Hecker M.;  
 RT "Identification and characterization of a new beta-glucoside  
 RT utilization system in *Bacillus subtilis*.";  
 RL J. Bacteriol. 179:496-506(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / BGSC1A1;  
 RX MEDLINE=97124196; PubMed=8969509;  
 RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,  
 RA Miwa Y., Fujita Y.;  
 RT "Sequencing of a 65 kb region of the *Bacillus subtilis* genome  
 RT containing the *lic* and *cel* loci, and creation of a 177 kb contig  
 RT covering the *gnt-sacXY* region.";  
 RL Microbiology 142:3113-3123(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*.";

RL Nature 390:249-256(1997).  
 RN [4]  
 RP MUTAGENESIS OF HIS-219; HIS-278; HIS-333; HIS-392 AND HIS-559.  
 RX MEDLINE=99369874; PubMed=10438772;  
 RA Tobisch S., Stuelke J., Hecker M.;  
 RT "Regulation of the lic operon of Bacillus subtilis and  
 RT characterization of potential phosphorylation sites of the LicR  
 RT regulator protein by site-directed mutagenesis.";   
 RL J. Bacteriol. 181:4995-5003(1999).  
 CC -!- FUNCTION: Positive regulator of the licABCH operon.  
 CC -!- MISCELLANEOUS: Might be activated by phosphorylation of certain  
 CC histidyl residues by HPr.  
 CC -!- SIMILARITY: Contains 1 PTS EIIA domain.  
 CC -!- SIMILARITY: Belongs to the transcriptional antiterminator bglG  
 CC family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z49992; CAA90284.1; -.  
 DR EMBL; D83026; BAA11742.1; -.  
 DR EMBL; Z99123; CAB15886.1; -.  
 DR PIR; H69651; H69651.  
 DR SubtiList; BGL1346; licR.  
 DR InterPro; IPR001550; BglG\_antitermin.  
 DR InterPro; IPR002178; PTS\_EIIA\_2.  
 DR Pfam; PF00874; PRD; 2.  
 DR Pfam; PF00359; PTS\_EIIA\_2; 1.  
 DR ProDom; PD001689; PTS\_EIIA\_2; 1.  
 DR PROSITE; PS00654; ANTITERMINATORS\_BGLG; FALSE\_NEG.  
 KW Transcription regulation; Activator; RNA-binding; Phosphorylation;  
 KW Complete proteome.  
 FT DOMAIN 499 638 EIIA DOMAIN.  
 FT MOD\_RES 219 219 PHOSPHORYLATION (POTENTIAL).  
 FT MOD\_RES 278 278 PHOSPHORYLATION (POTENTIAL).  
 FT MOD\_RES 333 333 PHOSPHORYLATION (POTENTIAL).  
 FT MOD\_RES 392 392 PHOSPHORYLATION (POTENTIAL).  
 FT MOD\_RES 559 559 PHOSPHORYLATION (POTENTIAL).  
 FT MUTAGEN 219 219 H->D: LOSS OF ACTIVITY.  
 FT MUTAGEN 219 219 H->A: RESIDUAL ACTIVITY.  
 FT MUTAGEN 278 278 H->E: LOSS OF ACTIVITY.  
 FT MUTAGEN 278 278 H->A: LOSS OF ACTIVITY.  
 FT MUTAGEN 333 333 H->A: LOSS OF ACTIVITY.  
 FT MUTAGEN 392 392 H->E: LOSS OF ACTIVITY.  
 FT MUTAGEN 392 392 H->I: RESIDUAL ACTIVITY.  
 FT MUTAGEN 559 559 H->G: INCREASE IN ACTIVITY.  
 SQ SEQUENCE 641 AA; 73315 MW; FCEF83BFC72A0168 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 641;  
 Best Local Similarity 85.7%; Pred. No. 99;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
| | | | |  
Db 505 LVFFQED 511

RESULT 32

BGAL\_LACDE

ID BGAL\_LACDE STANDARD; PRT; 1006 AA.  
AC P20043;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-MAY-1991 (Rel. 18, Last sequence update)  
DT 01-JUN-1994 (Rel. 29, Last annotation update)  
DE Beta-galactosidase (EC 3.2.1.23) (Lactase).  
GN LACZ.  
OS Lactobacillus delbrueckii (subsp. bulgaricus).  
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
OC Lactobacillus.  
OX NCBI\_TaxID=1585;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-29.  
RC STRAIN=B131;  
RX MEDLINE=89123132; PubMed=2492511;  
RA Schmidt B.F., Adams R.M., Requadt C., Power S., Mainzer S.E.;  
RT "Expression and nucleotide sequence of the Lactobacillus bulgaricus  
RT beta-galactosidase gene cloned in Escherichia coli.";  
RL J. Bacteriol. 171:625-635(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 11842;  
RX MEDLINE=91161509; PubMed=1705929;  
RA Leong-Morgenthaler P.M., Zwahlen M.C., Hottinger H.;  
RT "Lactose metabolism in Lactobacillus bulgaricus: analysis of the  
RT primary structure and expression of the genes involved.";  
RL J. Bacteriol. 173:1951-1957(1991).  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-  
CC galactose residues in beta-D-galactosides.  
CC -!- SUBUNIT: Monomer.  
CC -!- SIMILARITY: Belongs to family 2 of glycosyl hydrolases.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M23530; AAA25240.1; -.  
DR EMBL; M55068; AAA25244.1; -.  
DR PIR; A30093; A30093.  
DR HSSP; P00722; 1BGL.  
DR InterPro; IPR008979; Gal\_bind\_like.  
DR InterPro; IPR006101; Glyco\_hydro\_2.  
DR InterPro; IPR006102; Glyco\_hydro\_2Ig.  
DR InterPro; IPR006104; Glyco\_hydro\_2SB.  
DR InterPro; IPR006103; Glyco\_hydro\_2TIM.  
DR InterPro; IPR004200; Glyco\_hydro\_42C.

DR InterPro; IPR004199; Glyco\_hydro\_42N.  
 DR Pfam; PF02930; Bgal\_small\_C; 1.  
 DR Pfam; PF02929; Bgal\_small\_N; 1.  
 DR Pfam; PF00703; Glyco\_hydro\_2; 1.  
 DR Pfam; PF02836; Glyco\_hydro\_2\_C; 1.  
 DR Pfam; PF02837; Glyco\_hydro\_2\_N; 1.  
 DR PRINTS; PR00132; GLHYDRLASE2.  
 DR PROSITE; PS00719; GLYCOSYL\_HYDROL\_F2\_1; 1.  
 DR PROSITE; PS00608; GLYCOSYL\_HYDROL\_F2\_2; 1.  
 KW Hydrolase; Glycosidase.  
 FT INIT MET 0 0  
 FT ACT\_SITE 464 464 PROTON DONOR (BY SIMILARITY).  
 FT ACT\_SITE 531 531 NUCLEOPHILE (BY SIMILARITY).  
 FT CONFLICT 903 903 Y -> YK (IN REF. 2).  
 SQ SEQUENCE 1006 AA; 113915 MW; E5F81240EB64E769 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 1006;  
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 FFAEDF 8  
 |:||||  
 Db 73 FYAEDF 78

# RESULT 33

## UGG2\_HUMAN

ID UGG2\_HUMAN STANDARD; PRT; 1516 AA.  
 AC Q9NYU1; Q9UFC4;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE UDP-glucose:glycoprotein glucosyltransferase 2 precursor (EC 2.4.1.-)  
 DE (UDP--Glc:glycoprotein glucosyltransferase 2) (UGT 2) (HUGT2).  
 GN UGCGL2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Fetal liver;  
 RX MEDLINE=20160491; PubMed=10694380;  
 RA Arnold S.M., Fessler L.I., Fessler J.H., Kaufman R.J.;  
 RT "Two homologues encoding human UDP-glucose:glycoprotein  
 RT glucosyltransferase differ in mRNA expression and enzymatic  
 RT activity.";  
 RL Biochemistry 39:2149-2163(2000).  
 RN [2]  
 RP SEQUENCE OF 151-1516 FROM N.A.  
 RC TISSUE=Testis;  
 RA Bloecker H., Boecher M., Brandt P., Mewes H.-W., Gassenhuber J.,  
 RA Wiemann S.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: Selectively reglucosylates unfolded glycoproteins, thus  
 CC providing quality control for protein transport out of the ER.  
 CC Unfolded, denatured glycoproteins are substantially better

CC substrates for glucosylation by this enzyme than are the  
 CC corresponding native proteins. This protein and transient  
 CC glucosylation may be involved in monitoring and/or assisting the  
 CC folding and assembly of newly made glycoproteins, in order to  
 CC identify glycoproteins that need assistance in folding from  
 CC chaperones.  
 CC -!- COFACTOR: Requires calcium and manganese ions for activity.  
 CC -!- PATHWAY: Glycosylation.  
 CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.  
 CC -!- TISSUE SPECIFICITY: Expressed in pancreas, heart, brain, placenta,  
 CC lung, liver, skeletal muscle, and kidney.  
 CC -!- SIMILARITY: Belongs to the glycosyltransferase family 8.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AF227906; AAF66233.2; -.  
 DR EMBL; AL133051; CAB61378.1; -.  
 DR PIR; T42654; T42654.  
 DR Genew; HGNC:15664; UGCGL2.  
 DR MIM; 605898; -.  
 DR InterPro; IPR000886; ER\_target\_S.  
 DR InterPro; IPR002495; Glyco\_trans\_8.  
 DR Pfam; PF01501; Glyco\_transf\_8; 1.  
 DR PROSITE; PS00014; ER\_TARGET; 1.  
 KW Signal; Transferase; Glycosyltransferase; Endoplasmic reticulum;  
 KW Glycoprotein.  
 FT SIGNAL 1 27 POTENTIAL.  
 FT CHAIN 28 1516 UDP-GLUCOSE:GLYCOPROTEIN  
 FT GLUCOSYLTRANSFERASE 2.  
 FT CARBOHYD 256 256 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 920 920 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 950 950 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT SITE 1513 1516 PREVENT SECRETION FROM ER (POTENTIAL).  
 FT CONFLICT 323 323 T -> A (IN REF. 2).  
 FT CONFLICT 328 328 A -> S (IN REF. 2).  
 FT CONFLICT 821 821 T -> A (IN REF. 2).  
 FT CONFLICT 994 994 L -> M (IN REF. 2).  
 SQ SEQUENCE 1516 AA; 174760 MW; 31A7C0AFF29D1268 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 1516;  
 Best Local Similarity 83.3%; Pred. No. 2.3e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 FFAEDF 8  
 |:||||  
 Db 887 FYAEDF 892

RESULT 34  
 CA1E\_CHICK

ID CA1E\_CHICK STANDARD; PRT; 1888 AA.  
 AC P32018;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Collagen alpha 1(XIV) chain precursor (Undulin).  
 GN COL14A1.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=93185668; PubMed=8444186;  
 RA Waelchli C., Trueb J., Kessler B., Winterhalter K.H., Trueb B.;  
 RT "Complete primary structure of chicken collagen XIV.";  
 RL Eur. J. Biochem. 212:483-490(1993).  
 RN [2]  
 RP SEQUENCE OF 402-1549 FROM N.A.  
 RX MEDLINE=92339443; PubMed=1339349;  
 RA Trueb J., Trueb B.;  
 RT "Type XIV collagen is a variant of undulin.";  
 RL Eur. J. Biochem. 207:549-557(1992).  
 RN [3]  
 RP STRUCTURE BY NMR OF 1852-1885.  
 RX MEDLINE=99280705; PubMed=10350466;  
 RA Montserret R., Aubert-Foucher E., McLeish M.J., Hill J.M., Ficheux D.,  
 RA Jaquinod M., van der Rest M., Deleage G., Penin F.;  
 RT "Structural analysis of the heparin-binding site of the NC1 domain of  
 RT collagen XIV by CD and NMR.";  
 RL Biochemistry 38:6479-6488(1999).  
 CC -!- FUNCTION: An adhesive role by integrating collagen bundles. It is  
 CC probably associated with the surface of interstitial collagen  
 CC fibrils via COL1. The COL2 domain may then serve as a rigid arm  
 CC which sticks out from the fibril and protrudes the large N-  
 CC terminal globular domain into the extracellular space, where it  
 CC might interact with other matrix molecules or cell surface  
 CC receptors.  
 CC -!- SUBUNIT: Homotrimer (Probable).  
 CC -!- SUBCELLULAR LOCATION: Extracellular matrix.  
 CC -!- TISSUE SPECIFICITY: Wide tissue distribution; high presence in  
 CC dense connective tissue in skeletal muscle.  
 CC -!- PTM: Lysines at the third position of the tripeptide repeating  
 CC unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.  
 CC -!- PTM: Prolines at the third position of the tripeptide repeating  
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC -!- PTM: May contain numerous cysteine residues involved in inter- and  
 CC intramolecular disulfide bonding.  
 CC -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH  
 CC INTERRUPTED HELICES (FACIT) FAMILY.  
 CC -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.  
 CC -!- SIMILARITY: Contains 2 VWFA domains.  
 CC -!- SIMILARITY: Contains 6 fibronectin type III domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration



CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X70793; CAA50064.1; -.  
 DR EMBL; X70792; CAA50063.1; -.  
 DR EMBL; X66138; CAA46928.1; -.  
 DR PIR; A45974; A45974.  
 DR PIR; S78476; S78476.  
 DR PDB; 1B9P; 25-FEB-99.  
 DR PDB; 1B9Q; 25-FEB-99.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR008957; FN\_III-like.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR003129; TSPN.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF01391; Collagen; 4.  
 DR Pfam; PF00041; fn3; 8.  
 DR Pfam; PF02210; TSPN; 1.  
 DR Pfam; PF00092; vwa; 2.  
 DR PRINTS; PR00453; VWFADOMAIN.  
 DR SMART; SM00060; FN3; 6.  
 DR SMART; SM00210; TSPN; 1.  
 DR SMART; SM00327; VWA; 2.  
 DR PROSITE; PS50234; VWFA; 2.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Cell adhesion; Collagen; Glycoprotein; Signal; 3D-structure.  
 FT SIGNAL 1 28 POTENTIAL.  
 FT CHAIN 29 1888 COLLAGEN ALPHA 1(XIV) CHAIN.  
 FT DOMAIN 158 330 VWFA 1.  
 FT DOMAIN 441 529 FIBRONECTIN TYPE-III 1.  
 FT DOMAIN 531 622 FIBRONECTIN TYPE-III 2.  
 FT DOMAIN 625 728 FIBRONECTIN TYPE-III 3.  
 FT DOMAIN 739 821 FIBRONECTIN TYPE-III 4.  
 FT DOMAIN 829 922 FIBRONECTIN TYPE-III 5.  
 FT DOMAIN 952 1010 FIBRONECTIN TYPE-III 6.  
 FT DOMAIN 1042 1215 VWFA 2.  
 FT DOMAIN 1239 1444 TSP N-TERMINAL.  
 FT DOMAIN 1227 1468 NONHELICAL REGION (NC4).  
 FT DOMAIN 1469 1620 TRIPLE-HELICAL REGION 1.  
 FT DOMAIN 1664 1786 TRIPLE-HELICAL REGION 2.  
 FT DOMAIN 721 733 POLY-THR.  
 FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1398 1398 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT SITE 1489 1491 CELL ATTACHMENT SITE (POTENTIAL).  
 FT SITE 1617 1619 CELL ATTACHMENT SITE (POTENTIAL).  
 FT TURN 1855 1856  
 FT TURN 1859 1861  
 FT HELIX 1862 1880  
 FT TURN 1881 1882  
 SQ SEQUENCE 1888 AA; 202666 MW; 39915BB9F46DD873 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 1888;  
 Best Local Similarity 71.4%; Pred. No. 2.9e+02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
||| :||  
Db 1199 VFFVDDF 1205

RESULT 35

Y126\_PYRAB

ID Y126\_PYRAB STANDARD; PRT; 251 AA.  
AC Q9V2E8;  
DT 15-MAR-2004 (Rel. 43, Created)  
DT 15-MAR-2004 (Rel. 43, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Hypothetical UPF0286 protein PYRAB01260.  
GN PYRAB01260 OR PAB2263.  
OS Pyrococcus abyssi.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Pyrococcus.  
OX NCBI\_TaxID=29292;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GE5 / Orsay;  
RX MEDLINE=22511545; PubMed=12622808;  
RA Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,  
RA Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,  
RA Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;  
RT "An integrated analysis of the genome of the hyperthermophilic  
RT archaeon Pyrococcus abyssi.";  
RL Mol. Microbiol. 47:1495-1512(2003).  
CC -!- SIMILARITY: Belongs to the UPF0286 family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AJ248283; CAB49050.1; -.  
DR PIR; C75200; C75200.  
DR HAMAP; MF\_00722; -; 1.  
DR InterPro; IPR002793; DUF91.  
DR Pfam; PF01939; DUF91; 1.  
DR ProDom; PD013521; DUF91; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 251 AA; 28831 MW; 648BB6A7CE1E0EFC CRC64;

Query Match 70.7%; Score 29; DB 1; Length 251;  
Best Local Similarity 62.5%; Pred. No. 64;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
:|| |||:  
Db 111 VVFLAEDY 118

RESULT 36

NU1M\_DALCH

ID NU1M\_DALCH STANDARD; PRT; 310 AA.  
 AC 063623;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).  
 GN MTND1 OR ND1.  
 OS Dalbulus charlesi.  
 OG Mitochondrion.  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Paraneoptera; Hemiptera; Euhemiptera; Membracoidea;  
 OC Cicadellidae; Deltocephalinae; Dalbulus.  
 OX NCBI\_TaxID=74059;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Dietrich C.H., Fitzgerald S.J., Holmes J.L., Black W.C. IV,  
 RA Nault L.R.;  
 RT "Reassessment of Dalbulus leafhopper (Homoptera: Cicadellidae)  
 RT phylogeny based on mitochondrial DNA sequences.";  
 RL Ann. Entomol. Soc. Am. 91:590-597(1998).  
 CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
 CC -!- SIMILARITY: Belongs to the complex I subunit 1 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF051289; AAC05753.1; -.  
 DR InterPro; IPR001694; Resp\_NADH\_dh1.  
 DR Pfam; PF00146; NADHdh; 1.  
 DR PROSITE; PS00667; COMPLEX1\_ND1\_1; FALSE\_NEG.  
 DR PROSITE; PS00668; COMPLEX1\_ND1\_2; 1.  
 KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.  
 SQ SEQUENCE 310 AA; 36376 MW; FD59E413A881854B CRC64;

Query Match 70.7%; Score 29; DB 1; Length 310;  
 Best Local Similarity 50.0%; Pred. No. 79;  
 Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 :||| ||  
 Db 240 IIFFGSDF 247

RESULT 37

Y189\_RICPR

ID Y189\_RICPR STANDARD; PRT; 321 AA.  
 AC Q9ZDX5;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein RP189.  
 GN RP189.  
 OS Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
 OC Rickettsiaceae; Rickettsieae; Rickettsia.  
 OX NCBI\_TaxID=782;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Madrid E;  
 RX MEDLINE=99039499; PubMed=9823893;  
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
 RA Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
 RT "The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria.";  
 RL Nature 396:133-140(1998).  
 CC -!- SIMILARITY: SOME, TO A.AEOLICUS AQ\_1104.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AJ235270; CAA14655.1; -.  
 DR PIR; H71729; H71729.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 321 AA; 36653 MW; 3E5F47D104DD8A73 CRC64;

Query Match 70.7%; Score 29; DB 1; Length 321;  
 Best Local Similarity 71.4%; Pred. No. 82;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
 |:|||||  
 Db 179 LIFFAHD 185

#### RESULT 38

##### HST2\_YEAST

ID HST2\_YEAST STANDARD; PRT; 357 AA.  
 AC P53686;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE NAD-dependent deacetylase HST2 (EC 3.5.1.-) (Homologous to SIR2  
 DE protein 2).  
 GN HST2 OR YPL015C OR LPA2C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=S288c / YPH1;  
 RX MEDLINE=96101589; PubMed=7498786;  
 RA Brachmann C.B., Sherman J.M., Devine S.E., Cameron E.E., Pillus L.,  
 RA Boeke J.D.;  
 RT "The SIR2 gene family, conserved from bacteria to humans, functions  
 RT in silencing, cell cycle progression, and chromosome stability.";  
 RL Genes Dev. 9:2888-2902(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c / AB972;  
 RX MEDLINE=97313271; PubMed=9169875;  
 RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansorge W.,  
 RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,  
 RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,  
 RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,  
 RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,  
 RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,  
 RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.W.,  
 RA Hunicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,  
 RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,  
 RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,  
 RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,  
 RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,  
 RA Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,  
 RA Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,  
 RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,  
 RA Zhong W.W., Zollner A., Vo D.H., Hani J.;  
 RT "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XVI.";  
 RL Nature 387:103-105(1997).  
 CC -!- FUNCTION: NAD-dependent histone deacetylase, which could function  
 CC in telomeric silencing, cell cycle progression and chromosome  
 CC stability.  
 CC -!- CATALYTIC ACTIVITY: NAD(+) + an acetylprotein = nicotinamide + O-  
 CC acetyl-ADP-ribose + a protein.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -!- SIMILARITY: Belongs to the sirtuin family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U39063; AAA81035.1; -.  
 DR EMBL; U33335; AAB68090.1; -.  
 DR PIR; S59678; S59678.  
 DR GermOnline; 143997; -.  
 DR SGD; S0005936; HST2.  
 DR GO; GO:0005737; C:cytoplasm; IDA.  
 DR GO; GO:0017136; F:NAD-dependent histone deacetylase activity; IDA.  
 DR GO; GO:0006348; P:chromatin silencing at telomere; IGI.  
 DR InterPro; IPR003000; SIR2.  
 DR Pfam; PF02146; SIR2; 1.  
 DR PROSITE; PS50305; SIRTUIN; 1.  
 KW Hydrolase; NAD; Transcription regulation; Repressor; Nuclear protein;

KW Metal-binding; Zinc.  
 FT DOMAIN 13 286 DEACETYLASE SIRTUIN-TYPE.  
 FT ACT\_SITE 135 135 BY SIMILARITY.  
 FT METAL 143 143 ZINC (BY SIMILARITY).  
 FT METAL 146 146 ZINC (BY SIMILARITY).  
 FT METAL 170 170 ZINC (BY SIMILARITY).  
 FT METAL 173 173 ZINC (BY SIMILARITY).  
 SQ SEQUENCE 357 AA; 39979 MW; ED281E5B8241A4D0 CRC64;

Query Match 70.7%; Score 29; DB 1; Length 357;  
 Best Local Similarity 71.4%; Pred. No. 91;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
 :||| ||  
 Db 181 IVFFGED 187

# RESULT 39

## HYD2\_BRAJA

ID HYD2\_BRAJA STANDARD; PRT; 380 AA.  
 AC P31904;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hydrogenase expression/formation protein hypD2.  
 GN HYPD2 OR HYPD' OR HYPD OR BLL6928.  
 OS Bradyrhizobium japonicum.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Bradyrhizobiaceae; Bradyrhizobium.  
 OX NCBI\_TaxID=375;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=USDA 110;  
 RX MEDLINE=22484998; PubMed=12597275;  
 RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,  
 RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,  
 RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,  
 RA Tabata S.;  
 RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
 RT Bradyrhizobium japonicum USDA110.";  
 RL DNA Res. 9:189-197(2002).  
 RN [2]  
 RP SEQUENCE OF 34-380 FROM N.A.  
 RC STRAIN=CB1809;  
 RX MEDLINE=93287991; PubMed=8510650;  
 RA van Soom C., Verreth C., Sampaio M.J., Vanderleyden J.;  
 RT "Identification of a potential transcriptional regulator of  
 RT hydrogenase activity in free-living Bradyrhizobium japonicum  
 RT strains.";  
 RL Mol. Gen. Genet. 239:235-240(1993).  
 CC -!- SIMILARITY: Belongs to the hypD family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AP005960; BAC52193.1; -.  
DR EMBL; Z17373; CAA78988.1; -.  
DR PIR; S35229; S35229.  
DR InterPro; IPR002780; HypD.  
DR Pfam; PF01924; HypD; 1.  
KW Metal-binding; Complete proteome.  
FT METAL 64 64 POTENTIAL.  
FT METAL 67 67 POTENTIAL.  
FT CONFLICT 54 54 K -> A (IN REF. 2).  
FT CONFLICT 82 82 E -> M (IN REF. 2).  
FT CONFLICT 261 261 S -> T (IN REF. 2).  
SQ SEQUENCE 380 AA; 42477 MW; 2C416E80BE8C2939 CRC64;

Query Match 70.7%; Score 29; DB 1; Length 380;  
Best Local Similarity 83.3%; Pred. No. 97;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 FFAEDF 8  
| | | | : |  
Db 215 FFAEEF 220

#### RESULT 40

##### O94B\_DROME

ID O94B\_DROME STANDARD; PRT; 383 AA.  
AC Q9VCS8;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Putative odorant receptor 94b.  
GN OR94B OR CG6679.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).

CC -!- FUNCTION: Probable role in the odorant response, being an odorant  
 CC receptor.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -!- SIMILARITY: Belongs to family Dr-or of G-protein coupled  
 CC receptors.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AE003741; AAF56077.1; -.  
 DR FlyBase; FBgn0039034; Or94b.  
 DR InterPro; IPR004117; 7tm\_6.  
 DR Pfam; PF02949; 7tm\_6; 1.

KW Hypothetical protein; Transmembrane; G-protein coupled receptor;  
 KW Olfaction; Multigene family.

FT	DOMAIN	1	41	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	42	62	1 (POTENTIAL).
FT	DOMAIN	63	70	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	71	91	2 (POTENTIAL).
FT	DOMAIN	92	130	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	131	151	3 (POTENTIAL).
FT	DOMAIN	152	174	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	175	195	4 (POTENTIAL).
FT	DOMAIN	196	250	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	251	271	5 (POTENTIAL).
FT	DOMAIN	272	284	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	285	305	6 (POTENTIAL).



FT DOMAIN 306 358 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 359 379 7 (POTENTIAL).  
 FT DOMAIN 380 383 CYTOPLASMIC (POTENTIAL).  
 SQ SEQUENCE 383 AA; 45262 MW; 6475485085672892 CRC64;

Query Match 70.7%; Score 29; DB 1; Length 383;  
 Best Local Similarity 71.4%; Pred. No. 98;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 VFFAEDF 8  
 ||| ||:  
 Db 150 VFFQEDY 156

# RESULT 41

## HYPD\_RHILV

ID HYPD\_RHILV STANDARD; PRT; 385 AA.  
 AC P40598;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Hydrogenase expression/formation protein hypD.  
 GN HYPD.  
 OS Rhizobium leguminosarum (biovar viciae).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.  
 OX NCBI\_TaxID=387;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=128c53;  
 RX MEDLINE=93316844; PubMed=8326860;  
 RA Rey L., Murillo J., Hernando Y., Hidalgo E., Cabrera E., Imperial J.,  
 RA Ruiz-Argueeso T.;  
 RT "Molecular analysis of a microaerobically induced operon required for  
 RT hydrogenase synthesis in Rhizobium leguminosarum biovar viciae."  
 RL Mol. Microbiol. 8:471-481(1993).  
 CC -!- SIMILARITY: Belongs to the hypD family.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X52974; CAA37163.1; -.  
 DR PIR; S32877; S32877.  
 DR InterPro; IPR002780; HypD.  
 DR Pfam; PF01924; HypD; 1.  
 DR TIGRFAMs; TIGR00075; hypD; 1.  
 KW Metal-binding.  
 FT METAL 64 64 POTENTIAL.  
 FT METAL 67 67 POTENTIAL.  
 SQ SEQUENCE 385 AA; 42028 MW; 0B15216C15665B12 CRC64;

Query Match 70.7%; Score 29; DB 1; Length 385;

Best Local Similarity 83.3%; Pred. No. 98;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 FFAEDF 8  
||||:|  
Db 215 FFAEEF 220

RESULT 42

SYH\_MYCPU

ID SYH\_MYCPU STANDARD; PRT; 420 AA.  
AC Q98QM8;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Histidyl-tRNA synthetase (EC 6.1.1.21) (Histidine--tRNA ligase)  
DE (HisRS).  
GN HISS OR MYPY\_3330.  
OS Mycoplasma pulmonis.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2107;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=UAB CTIP;  
RX MEDLINE=21267165; PubMed=11353084;  
RA Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,  
RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,  
RA Blanchard A.;  
RT "The complete genome sequence of the murine respiratory pathogen  
RT Mycoplasma pulmonis.";  
RL Nucleic Acids Res. 29:2145-2153(2001).  
CC -!- CATALYTIC ACTIVITY: ATP + L-histidine + tRNA(His) = AMP +  
CC diphosphate + L-histidyl-tRNA(His).  
CC -!- SUBUNIT: Homodimer (By similarity).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AL445564; CAC13506.1; -.  
DR PIR; E90553; E90553.  
DR MypuList; MYPY\_3330; -.  
DR HAMAP; MF\_00127; -; 1.  
DR InterPro; IPR004516; HisS.  
DR InterPro; IPR002314; tRNA-synt\_2b.  
DR InterPro; IPR006195; tRNA\_ligase\_II.  
DR Pfam; PF00587; tRNA-synt\_2b; 1.  
DR TIGRFAMs; TIGR00442; hisS; 1.  
DR PROSITE; PS50862; AA\_TRNA\_LIGASE\_II; FALSE\_NEG.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KW Complete proteome.

SQ SEQUENCE 420 AA; 49308 MW; 951B094769B34915 CRC64;

Query Match 70.7%; Score 29; DB 1; Length 420;  
Best Local Similarity 50.0%; Pred. No. 1.1e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
::|||::|  
Db 141 IIFFAKNF 148

RESULT 43

CLN3\_CANFA

ID CLN3\_CANFA STANDARD; PRT; 438 AA.  
AC Q29611;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE CLN3 protein (Battenin).  
GN CLN3.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98250317; PubMed=9590435;  
RA Shibuya H., Liu P.-C., Katz M.L., Siakotos A.N., Nonneman D.J.,  
RA Johnson G.S.;  
RT "Coding sequence and exon/intron organization of the canine CLN3  
RT (Batten disease) gene and its exclusion as the locus for  
RT ceroid-lipofuscinosis in English setter dogs.";  
RL J. Neurosci. Res. 52:268-275(1998).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Lysosomal (By  
CC similarity).  
CC -!- SIMILARITY: Belongs to the battenin family.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; L76281; AAB05546.1; -.  
DR EMBL; L76282; AAB05547.1; -.  
DR EMBL; AF033661; AAB86972.1; -.  
DR EMBL; AF033656; AAB86972.1; JOINED.  
DR EMBL; AF033657; AAB86972.1; JOINED.  
DR EMBL; AF033658; AAB86972.1; JOINED.  
DR EMBL; AF033659; AAB86972.1; JOINED.  
DR EMBL; AF033660; AAB86972.1; JOINED.  
DR InterPro; IPR003492; Battenin.  
DR Pfam; PF02487; CLN3; 1.  
DR PRINTS; PR01315; BATTENIN.  
KW Transmembrane; Lysosome; Glycoprotein.

FT	TRANSMEM	38	58	POTENTIAL.
FT	TRANSMEM	99	119	POTENTIAL.
FT	TRANSMEM	128	148	POTENTIAL.
FT	TRANSMEM	180	200	POTENTIAL.
FT	TRANSMEM	212	232	POTENTIAL.
FT	TRANSMEM	278	298	POTENTIAL.
FT	TRANSMEM	358	378	POTENTIAL.
FT	TRANSMEM	407	427	POTENTIAL.
SQ	SEQUENCE	438 AA; 47933 MW; 7964662FCDB33879 CRC64;		

Query Match 70.7%; Score 29; DB 1; Length 438;  
 Best Local Similarity 75.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 ||:|||| |  
 Db 290 LVYFAEYF 297

RESULT 44

CLN3\_MOUSE

ID CLN3\_MOUSE STANDARD; PRT; 438 AA.  
 AC Q61124; P70400;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE CLN3 protein (Battenin).  
 GN CLN3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Teratocarcinoma;  
 RX MEDLINE=97001171; PubMed=8812504;  
 RA Lee R.L., Johnson K.R., Lerner T.J.;  
 RT "Isolation and chromosomal mapping of a mouse homolog of the Batten  
 RT disease gene CLN3."  
 RL Genomics 35:617-619(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Liver;  
 RX MEDLINE=97295723; PubMed=9151313;  
 RA Taschner P.E.M., de Vos N., Breuning M.H.;  
 RT "Cross-species homology of the CLN3 gene."  
 RL Neuropediatrics 28:18-20(1997).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Lysosomal (By  
 CC similarity).  
 CC -!- SIMILARITY: Belongs to the battenin family.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch).

CC

DR EMBL; U47106; AAC52957.1; -.

DR EMBL; U68064; AAB07595.1; -.

DR MGD; MGI:107537; Cln3.

DR InterPro; IPR003492; Battenin.

DR Pfam; PF02487; CLN3; 1.

DR PRINTS; PR01315; BATTENIN.

KW Transmembrane; Lysosome; Glycoprotein.

FT TRANSMEM 38 58 POTENTIAL.

FT TRANSMEM 99 119 POTENTIAL.

FT TRANSMEM 128 148 POTENTIAL.

FT TRANSMEM 187 207 POTENTIAL.

FT TRANSMEM 212 232 POTENTIAL.

FT TRANSMEM 278 298 POTENTIAL.

FT TRANSMEM 358 378 POTENTIAL.

FT TRANSMEM 407 427 POTENTIAL.

FT CONFLICT 79 79 R -> P (IN REF. 2).

FT CONFLICT 120 120 H -> Y (IN REF. 2).

FT CONFLICT 237 237 W -> L (IN REF. 2).

FT CONFLICT 317 318 HE -> QQ (IN REF. 2).

FT CONFLICT 380 380 H -> Y (IN REF. 2).

SQ SEQUENCE 438 AA; 47773 MW; 0983C604E6959A2C CRC64;

Query Match 70.7%; Score 29; DB 1; Length 438;

Best Local Similarity 75.0%; Pred. No. 1.1e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8

||:||||

Db 290 LVYFAEYF 297

#### RESULT 45

CH62\_CHLPN

ID CH62\_CHLPN STANDARD; PRT; 526 AA.

AC Q9Z7C9; Q9JS18;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE 60 kDa chaperonin 2 (Protein Cpn60 2) (groEL protein 2).

GN GROL2 OR GROEL2 OR CPN0777 OR CP1095 OR CPB0805.

OS Chlamydia pneumoniae (Chlamydophila pneumoniae).

OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.

OX NCBI\_TaxID=83558;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CWL029;

RX MEDLINE=99206606; PubMed=10192388;

RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,

RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;

RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";

RL Nat. Genet. 21:385-389(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=AR39;

RX MEDLINE=20150255; PubMed=10684935;

RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
 RT pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CWL029 from USA.";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TW-183;  
 RA Geng M.M., Schuhmacher A., Muehldorfer I., Bensch K.W., Schaefer K.P.,  
 RA Schneider S., Pohl T., Essig A., Marre R., Melchers K.;  
 RT "The genome sequence of Chlamydia pneumoniae TW183 and comparison with  
 RT other Chlamydia strains based on whole genome sequence analysis.";  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: Prevents misfolding and promotes the refolding and  
 CC proper assembly of unfolded polypeptides generated under stress  
 CC conditions (By similarity).  
 CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of  
 CC 7 subunits (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE001659; AAD18915.1; -.  
 DR EMBL; AE002265; AAF38865.1; -.  
 DR EMBL; AP002547; BAA98985.1; -.  
 DR EMBL; AE017159; AAP98734.1; -.  
 DR PIR; D72036; D72036.  
 DR PIR; F81504; F81504.  
 DR PIR; G86587; G86587.  
 DR HSSP; P06139; 1GRL.  
 DR TIGR; CP1095; -.  
 DR HAMAP; MF\_00600; atypical; 1.  
 DR InterPro; IPR001844; Chaperonin\_Cpn60.  
 DR InterPro; IPR002423; Cpn60/TCP-1.  
 DR InterPro; IPR008950; GroEL-ATPase.  
 DR Pfam; PF00118; cpn60\_TCP1; 1.  
 DR PRINTS; PR00298; CHAPERONIN60.  
 DR PRINTS; PR00304; TCOMPLEXTCP1.

DR PROSITE; PS00296; CHAPERONINS\_CPN60; FALSE\_NEG.  
KW Chaperone; ATP-binding; Complete proteome.  
FT VARIANT 353 353 G -> R (in strains CWL029 and TW-183).  
SQ SEQUENCE 526 AA; 56825 MW; 60AAEFABB7D56238 CRC64;

Query Match 70.7%; Score 29; DB 1; Length 526;  
Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 LVFFAEDF 8  
|| ||||  
Db 245 LVIIAEDF 252

#### RESULT 46

##### UVRB\_STRP3

ID UVRB\_STRP3 STANDARD; PRT; 663 AA.  
AC Q8K743;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE UvrABC system protein B (UvrB protein) (Excinuclease ABC subunit B).  
GN UVRB OR SPYM3\_0996 OR SPS0862.  
OS Streptococcus pyogenes (serotype M3).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=198466;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MGAS315 / Serotype M3;  
RX MEDLINE=22133808; PubMed=12122206;  
RA Beres S.B., Sylva G.L., Barbian K.D., Lei B., Hoff J.S.,  
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,  
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,  
RA Schlievert P.M., Musser J.M.;  
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:  
RT phage-encoded toxins, the high-virulence phenotype, and clone  
RT emergence.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SSI-1 / Serotype M3;  
RX MEDLINE=22683278; PubMed=12799345;  
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,  
RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,  
RA Hayashi H., Hattori M., Hamada S.;  
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a  
RT large-scale genomic rearrangement in invasive strains and new insights  
RT into phage evolution.";  
RL Genome Res. 13:1042-1055(2003).  
CC -!- FUNCTION: The UvrABC repair system catalyzes the recognition and  
CC processing of DNA lesions. A damage recognition complex composed  
CC of 2 uvrA and 2 uvrB subunits scans DNA for abnormalities. Upon  
CC binding of the uvrA(2)B(2) complex to a putative damaged site, the  
CC DNA wraps around one uvrB monomer. DNA wrap is dependent on ATP  
CC binding by uvrB and probably causes local melting of the DNA  
CC helix, facilitating insertion of uvrB beta-hairpin between the DNA

```

CC      strands. Then uvrB probes one DNA strand for the presence of a
CC      lesion. If a lesion is found the uvrA subunits dissociate and the
CC      uvrB-DNA preincision complex is formed. This complex is
CC      subsequently bound by uvrC and the second uvrB is released. If no
CC      lesion is found, the DNA wraps around the other uvrB subunit that
CC      will check the other stand for damage (By similarity).
CC      -!- SUBUNIT: Forms a heterotetramer with uvrA during the search for
CC      lesions. Interacts with uvrC in an incision complex (By
CC      similarity).
CC      -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC      -!- DOMAIN: The beta-hairpin motif is involved in DNA binding (By
CC      similarity).
CC      -!- SIMILARITY: Belongs to the uvrB family.
CC      -!- SIMILARITY: Contains 1 UVR domain.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AE014156; AAM79603.1; -.
DR      EMBL; AP005143; BAC63957.1; -.
DR      HAMAP; MF_00204; -; 1.
DR      InterPro; IPR001410; DEAD.
DR      InterPro; IPR001650; Helicase_C.
DR      InterPro; IPR004807; UvrB.
DR      InterPro; IPR001943; UvrB/C.
DR      Pfam; PF00271; helicase_C; 1.
DR      Pfam; PF02151; UVR; 1.
DR      SMART; SM00487; DEXDc; 1.
DR      SMART; SM00490; HELICc; 1.
DR      TIGRFAMs; TIGR00631; uvrB; 1.
DR      PROSITE; PS50151; UVR; 1.
KW      SOS response; Excision nuclease; DNA repair; DNA recombination;
KW      DNA excision; ATP-binding; Complete proteome.
FT      NP_BIND      44      51      ATP (POTENTIAL).
FT      DOMAIN      97      120      BETA-HAIRPIN.
FT      DOMAIN      627     662      UVR.
SQ      SEQUENCE    663 AA;  75664 MW;  E4DA53524A14B0ED CRC64;

```

```

Query Match      70.7%;  Score 29;  DB 1;  Length 663;
Best Local Similarity 75.0%;  Pred. No. 1.7e+02;
Matches      6;  Conservative      0;  Mismatches      2;  Indels      0;  Gaps      0;

```

```

Qy      1 LVFFAEDF 8
        | || |||
Db      331 LDFFPEDF 338

```

```

RESULT 47
UVRB_STRP8
ID      UVRB_STRP8      STANDARD;      PRT;      663 AA.
AC      Q8P0J7;
DT      28-FEB-2003 (Rel. 41, Created)

```



DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE UvrABC system protein B (UvrB protein) (Excinuclease ABC subunit B).  
 GN UVRB OR SPYM18\_1326.  
 OS Streptococcus pyogenes (serotype M18).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=186103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MGAS8232 / Serotype M18;  
 RX MEDLINE=21927593; PubMed=11917108;  
 RA Smoot J.C., Barbian K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,  
 RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,  
 RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,  
 RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;  
 RT "Genome sequence and comparative microarray analysis of serotype M18  
 RT group A Streptococcus strains associated with acute rheumatic fever  
 RT outbreaks.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).  
 CC -!- FUNCTION: The UvrABC repair system catalyzes the recognition and  
 CC processing of DNA lesions. A damage recognition complex composed  
 CC of 2 uvrA and 2 uvrB subunits scans DNA for abnormalities. Upon  
 CC binding of the uvrA(2)B(2) complex to a putative damaged site, the  
 CC DNA wraps around one uvrB monomer. DNA wrap is dependent on ATP  
 CC binding by uvrB and probably causes local melting of the DNA  
 CC helix, facilitating insertion of uvrB beta-hairpin between the DNA  
 CC strands. Then uvrB probes one DNA strand for the presence of a  
 CC lesion. If a lesion is found the uvrA subunits dissociate and the  
 CC uvrB-DNA preincision complex is formed. This complex is  
 CC subsequently bound by uvrC and the second uvrB is released. If no  
 CC lesion is found, the DNA wraps around the other uvrB subunit that  
 CC will check the other stand for damage (By similarity).  
 CC -!- SUBUNIT: Forms a heterotetramer with uvrA during the search for  
 CC lesions. Interacts with uvrC in an incision complex (By  
 CC similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- DOMAIN: The beta-hairpin motif is involved in DNA binding (By  
 CC similarity).  
 CC -!- SIMILARITY: Belongs to the uvrB family.  
 CC -!- SIMILARITY: Contains 1 UVR domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE010054; AAL97930.1; -.  
 DR HAMAP; MF\_00204; -, 1.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004807; UvrB.  
 DR InterPro; IPR001943; UvrB/C.  
 DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF02151; UVR; 1.  
 DR TIGRFAMs; TIGR00631; uvrB; 1.  
 DR PROSITE; PS50151; UVR; 1.  
 KW SOS response; Excision nuclease; DNA repair; DNA recombination;  
 KW DNA excision; ATP-binding; Complete proteome.  
 FT NP\_BIND 44 51 ATP (POTENTIAL).  
 FT DOMAIN 97 120 BETA-HAIRPIN.  
 FT DOMAIN 627 662 UVR.  
 SQ SEQUENCE 663 AA; 75697 MW; D2EC5BFEC34B245 CRC64;

Query Match 70.7%; Score 29; DB 1; Length 663;  
 Best Local Similarity 75.0%; Pred. No. 1.7e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 | || |||  
 Db 331 LDFFPEDF 338

# RESULT 48

## UVRB\_STRPY

ID UVRB\_STRPY STANDARD; PRT; 663 AA.  
 AC Q99ZA5;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE UvrABC system protein B (UvrB protein) (Excinuclease ABC subunit B).  
 GN UVRB OR SPY1314.  
 OS Streptococcus pyogenes.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1314;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SF370 / ATCC 700294 / Serotype M1;  
 RX MEDLINE=21192684; PubMed=11296296;  
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
 RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,  
 RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,  
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;  
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
 CC -!- FUNCTION: The UvrABC repair system catalyzes the recognition and  
 CC processing of DNA lesions. A damage recognition complex composed  
 CC of 2 uvrA and 2 uvrB subunits scans DNA for abnormalities. Upon  
 CC binding of the uvrA(2)B(2) complex to a putative damaged site, the  
 CC DNA wraps around one uvrB monomer. DNA wrap is dependent on ATP  
 CC binding by uvrB and probably causes local melting of the DNA  
 CC helix, facilitating insertion of uvrB beta-hairpin between the DNA  
 CC strands. Then uvrB probes one DNA strand for the presence of a  
 CC lesion. If a lesion is found the uvrA subunits dissociate and the  
 CC uvrB-DNA preincision complex is formed. This complex is  
 CC subsequently bound by uvrC and the second uvrB is released. If no  
 CC lesion is found, the DNA wraps around the other uvrB subunit that  
 CC will check the other stand for damage (By similarity).  
 CC -!- SUBUNIT: Forms a heterotetramer with uvrA during the search for  
 CC lesions. Interacts with uvrC in an incision complex (By

```

CC      similarity).
CC      -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC      -!- DOMAIN: The beta-hairpin motif is involved in DNA binding (By
CC      similarity).
CC      -!- SIMILARITY: Belongs to the uvrB family.
CC      -!- SIMILARITY: Contains 1 UVR domain.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AE006570; AAK34157.1; -.
DR      HSSP; P56981; 1D9X.
DR      HAMAP; MF_00204; -; 1.
DR      InterPro; IPR001410; DEAD.
DR      InterPro; IPR001650; Helicase_C.
DR      InterPro; IPR004807; UvrB.
DR      InterPro; IPR001943; UvrB/C.
DR      Pfam; PF00271; helicase_C; 1.
DR      Pfam; PF02151; UVR; 1.
DR      SMART; SM00487; DEXDc; 1.
DR      SMART; SM00490; HELICc; 1.
DR      TIGRFAMs; TIGR00631; uvrB; 1.
DR      PROSITE; PS50151; UVR; 1.
KW      SOS response; Excision nuclease; DNA repair; DNA recombination;
KW      DNA excision; ATP-binding; Complete proteome.
FT      NP_BIND      44      51      ATP (POTENTIAL).
FT      DOMAIN      97      120      BETA-HAIRPIN.
FT      DOMAIN      627     662      UVR.
SQ      SEQUENCE     663 AA;  75654 MW;  999CB73C0D49C53A CRC64;

```

```

Query Match          70.7%; Score 29; DB 1; Length 663;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches      6; Conservative      0; Mismatches      2; Indels      0; Gaps      0;

```

```

Qy      1 LVFFAEDF 8
        | || |||
Db      331 LDFFPEDF 338

```

#### RESULT 49

##### PTR3\_YEAST

```

ID      PTR3_YEAST      STANDARD;      PRT;      678 AA.
AC      P43606;
DT      01-NOV-1995 (Rel. 32, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      PTR3 protein (SSY3 protein).
GN      PTR3 OR SSY3 OR YFR029W.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX      NCBI_TaxID=4932;

```

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c / AB972;  
 RX MEDLINE=95400292; PubMed=7670463;  
 RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,  
 RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,  
 RA Yamazaki M., Tashiro H., Eki T.;  
 RT "Analysis of the nucleotide sequence of chromosome VI from  
 RT *Saccharomyces cerevisiae*.";  
 RL Nat. Genet. 10:261-268(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c / AB972;  
 RX MEDLINE=96287654; PubMed=8686381;  
 RA Eki T., Naitou M., Hagiwara H., Abe M., Ozawa M., Sasanuma S.-I.,  
 RA Sasanuma M., Tsuchiya Y., Shibata T., Watanabe K., Ono A.,  
 RA Yamazaki M.-A., Tashiro H., Hanaoka F., Murakami Y.;  
 RT "Fifteen open reading frames in a 30.8 kb region of the right arm of  
 RT chromosome VI from *Saccharomyces cerevisiae*.";  
 RL Yeast 12:177-190(1996).  
 RN [3]  
 RP IDENTIFICATION.  
 RX MEDLINE=98144794; PubMed=9483800;  
 RA Joergensen M.U., Bruun M.B., Didion T., Kielland-Brandt M.C.;  
 RT "Mutations in five loci affecting GAP1-independent uptake of neutral  
 RT amino acids in yeast.";  
 RL Yeast 14:103-114(1998).  
 CC -!- FUNCTION: Regulator of the expression of the transporters BAP2,  
 CC GAP1 and PTR2.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; D50617; BAA09268.1; -.  
 DR PIR; S56284; S56284.  
 DR GermOnline; 140183; -.  
 DR SGD; S0001925; PTR3.  
 DR GO; GO:0019897; C:extrinsic to plasma membrane; IDA.  
 DR GO; GO:0016597; F:amino acid binding; IGI.  
 DR GO; GO:0007606; P:chemosensory perception; IGI.  
 SQ SEQUENCE 678 AA; 76286 MW; BF2B64842D669D6B CRC64;

Query Match 70.7%; Score 29; DB 1; Length 678;  
 Best Local Similarity 83.3%; Pred. No. 1.7e+02;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 FFAEDF 8  
 || |||  
 Db 654 FFVEDF 659

## KCB2\_CANFA

ID KCB2\_CANFA STANDARD; PRT; 809 AA.  
AC Q95167;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Potassium voltage-gated channel subfamily B member 2 (Potassium  
DE channel Kv2.2) (Fragment).  
GN KCNB2.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98275219; PubMed=9612272;  
RA Schmalz F., Kinsella J., Koh S.D., Vogalis F., Schneider A.,  
RA Flynn E.R., Kenyon J.L., Horowitz B.;  
RT "Molecular identification of a component of delayed rectifier current  
RT in gastrointestinal smooth muscles.";  
RL Am. J. Physiol. 274:G901-G911(1998).  
RN [2]  
RP CONCEPTUAL TRANSLATION.  
RA Sundaram S.;  
RL Unpublished observations (OCT-2002).  
CC -!- FUNCTION: Mediates the voltage-dependent potassium ion  
CC permeability of excitable membranes. Channels open or close in  
CC response to the voltage difference across the membrane, letting K+  
CC ions pass in accordance with their electrochemical gradient.  
CC -!- SUBUNIT: Heteromultimer with KCNS1, KCNS2 and KCNS3 (By  
CC similarity).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- DOMAIN: The segment S4 is probably the voltage-sensor and is  
CC characterized by a series of positively charged amino acids at  
CC every third position.  
CC -!- DOMAIN: The tail may be important in modulation of channel  
CC activity and/or targeting of the channel to specific subcellular  
CC compartments.  
CC -!- PTM: Phosphorylated (Potential).  
CC -!- SIMILARITY: Belongs to the potassium channel family. B (Shab)  
CC subfamily.  
CC -!- CAUTION: This is a conceptual translation, a frameshift was  
CC introduced in position 768 to restore the similarity with the  
CC other orthologs.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U69963; AAB08432.1; ALT\_FRAME.  
DR HSSP; Q54397; 1BL8.  
DR InterPro; IPR000210; BTB\_POZ.  
DR InterPro; IPR005821; Ion\_trans.

DR InterPro; IPR001622; K+channel\_pore.  
 DR InterPro; IPR003091; K\_channel.  
 DR InterPro; IPR003131; K\_tetra.  
 DR InterPro; IPR005826; Kv22channel.  
 DR InterPro; IPR003968; Kv\_channel.  
 DR InterPro; IPR005820; M+channel\_nlg.  
 DR InterPro; IPR003973; Shab\_channel.  
 DR Pfam; PF00520; ion\_trans; 1.  
 DR Pfam; PF02214; K\_tetra; 1.  
 DR Pfam; PF03521; Kv2channel; 1.  
 DR PRINTS; PR00169; KCHANNEL.  
 DR PRINTS; PR01515; KV22CHANNEL.  
 DR PRINTS; PR01491; KVCHANNEL.  
 DR PRINTS; PR01495; SHABCHANNEL.  
 DR SMART; SM00225; BTB; 1.  
 KW Transport; Ion transport; Ionic channel; Voltage-gated channel;  
 KW Potassium channel; Potassium; Potassium transport; Transmembrane;  
 KW Phosphorylation; Glycoprotein; Multigene family.  
 FT DOMAIN 1 190 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 191 212 SEGMENT S1 (POTENTIAL).  
 FT TRANSMEM 233 254 SEGMENT S2 (POTENTIAL).  
 FT DOMAIN 255 264 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 265 286 SEGMENT S3 (POTENTIAL).  
 FT TRANSMEM 299 320 SEGMENT S4 (POTENTIAL).  
 FT DOMAIN 321 334 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 335 356 SEGMENT S5 (POTENTIAL).  
 FT DOMAIN 369 389 SEGMENT H5 (PORE-FORMING) (POTENTIAL).  
 FT TRANSMEM 397 418 SEGMENT S6 (POTENTIAL).  
 FT DOMAIN 419 >809 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT NON\_TER 809 809  
 SQ SEQUENCE 809 AA; 90602 MW; E1FF56334F3DA9C8 CRC64;

Query Match 70.7%; Score 29; DB 1; Length 809;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVFFAE 6  
 |||||  
 Db 352 LVFFAE 357

RESULT 51  
 KCBl\_MOUSE  
 ID KCBl\_MOUSE STANDARD; PRT; 857 AA.  
 AC Q03717;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Potassium voltage-gated channel subfamily B member 1 (Potassium  
 DE channel Kv2.1) (mShab).  
 GN KCNB1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=91162315; PubMed=2002364;  
 RA Pak M.D., Covarrubias M., Ratcliffe A., Salkoff L.;  
 RT "A mouse brain homolog of the Drosophila Shab K<sup>+</sup> channel with  
 RT conserved delayed-rectifier properties.";  
 RL J. Neurosci. 11:869-880(1991).  
 CC -!- FUNCTION: Mediates the voltage-dependent potassium ion  
 CC permeability of excitable membranes. Channels open or close in  
 CC response to the voltage difference across the membrane, letting K<sup>+</sup>  
 CC ions pass in accordance with their electrochemical gradient.  
 CC -!- SUBUNIT: Heteromultimer with KCNG2, KCNG3, KCNG4, KCNS1, KCNS2,  
 CC KCNS3 and KCNV2 (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- DOMAIN: The segment S4 is probably the voltage-sensor and is  
 CC characterized by a series of positively charged amino acids at  
 CC every third position.  
 CC -!- DOMAIN: The tail may be important in modulation of channel  
 CC activity and/or targeting of the channel to specific subcellular  
 CC compartments.  
 CC -!- PTM: Phosphorylated on serine residues in the C-terminal  
 CC cytoplasmic tail (By similarity).  
 CC -!- MISCELLANEOUS: Inhibited by tetraethylammonium chloride (TEA).  
 CC -!- SIMILARITY: Belongs to the potassium channel family. B (Shab)  
 CC subfamily.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; M64228; AAA40112.1; -.  
 DR PIR; I56529; I56529.  
 DR HSSP; Q54397; 1BL8.  
 DR MGD; MGI:96666; Kcnb1.  
 DR InterPro; IPR000210; BTB\_POZ.  
 DR InterPro; IPR005821; Ion\_trans.  
 DR InterPro; IPR001622; K<sup>+</sup>channel\_pore.  
 DR InterPro; IPR003091; K\_channel.  
 DR InterPro; IPR003131; K\_tetra.  
 DR InterPro; IPR004350; Kv21channel.  
 DR InterPro; IPR003968; Kv\_channel.  
 DR InterPro; IPR005820; M<sup>+</sup>channel\_nlg.  
 DR InterPro; IPR003973; Shab\_channel.  
 DR Pfam; PF00520; ion\_trans; 1.  
 DR Pfam; PF02214; K\_tetra; 1.  
 DR Pfam; PF03521; Kv2channel; 1.  
 DR PRINTS; PR00169; KCHANNEL.  
 DR PRINTS; PR01514; KV21CHANNEL.  
 DR PRINTS; PR01491; KVCHANNEL.  
 DR PRINTS; PR01495; SHABCHANNEL.  
 DR SMART; SM00225; BTB; 1.  
 KW Transport; Ion transport; Ionic channel; Voltage-gated channel;  
 KW Potassium channel; Potassium; Potassium transport; Transmembrane;

KW Phosphorylation; Multigene family.

FT	DOMAIN	1	185	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	186	206	SEGMENT S1 (POTENTIAL).
FT	TRANSMEM	229	249	SEGMENT S2 (POTENTIAL).
FT	DOMAIN	250	259	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	260	280	SEGMENT S3 (POTENTIAL).
FT	TRANSMEM	295	315	SEGMENT S4 (POTENTIAL).
FT	DOMAIN	316	330	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	331	351	SEGMENT S5 (POTENTIAL).
FT	DOMAIN	365	385	SEGMENT H5 (PORE-FORMING) (POTENTIAL).
FT	TRANSMEM	393	413	SEGMENT S6 (POTENTIAL).
FT	DOMAIN	414	857	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	517	520	POLY-SER.

SQ SEQUENCE 857 AA; 95692 MW; 772F1C42C25EF3A5 CRC64;

Query Match 70.7%; Score 29; DB 1; Length 857;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAE 6  
 |||||  
 Db 348 LVFFAE 353

# RESULT 52

## KCB1\_RAT

ID KCB1\_RAT STANDARD; PRT; 857 AA.  
 AC P15387;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Potassium voltage-gated channel subfamily B member 1 (Potassium channel Kv2.1) (DRK1).  
 GN KCNB1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE OF 4-857 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=89365157; PubMed=2770868;  
 RA Frech G.C., Vandongen A.M.J., Schuster G., Brown A.M., Joho R.H.;  
 RT "A novel potassium channel with delayed rectifier properties isolated from rat brain by expression cloning."  
 RL Nature 340:642-645(1989).  
 RN [2]  
 RP REVISIONS.  
 RA Frech G.C.;  
 RL Submitted (FEB-1990) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1-575 FROM N.A.  
 RX MEDLINE=92156897; PubMed=1740690;  
 RA Drewe J.A., Verma S., Frech G.C., Joho R.H.;  
 RT "Distinct spatial and temporal expression patterns of K+ channel mRNAs from different subfamilies."  
 RL J. Neurosci. 12:538-548(1992).



RN [4]  
 RP PHOSPHORYLATION, AND MUTAGENESIS OF SER-444 AND SER-496.  
 RX MEDLINE=98016256; PubMed=9351973;  
 RA Murakoshi H., Shi G., Scannevin R.H., Trimmer J.S.;  
 RT "Phosphorylation of the Kv2.1 K<sup>+</sup> channel alters voltage-dependent  
 RT activation.";  
 RL Mol. Pharmacol. 52:821-828(1997).  
 CC -!- FUNCTION: Mediates the voltage-dependent potassium ion  
 CC permeability of excitable membranes. Channels open or close in  
 CC response to the voltage difference across the membrane, letting K<sup>+</sup>  
 CC ions pass in accordance with their electrochemical gradient.  
 CC -!- SUBUNIT: Heteromultimer with KCNG2, KCNG3, KCNG4, KCNS1, KCNS2,  
 CC KCNS3 and KCNV2 (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- TISSUE SPECIFICITY: In the brain, the greatest density occurs in  
 CC the cerebral cortex, followed by the hippocampus, cerebellum, and  
 CC olfactory bulb. In peripheral tissues it is most prominent in  
 CC retina and kidney. Also present in cardiac muscle tissue of the  
 CC atrium and ventricle and in skeletal muscle.  
 CC -!- DOMAIN: The segment S4 is probably the voltage-sensor and is  
 CC characterized by a series of positively charged amino acids at  
 CC every third position.  
 CC -!- DOMAIN: The tail may be important in modulation of channel  
 CC activity and/or targeting of the channel to specific subcellular  
 CC compartments.  
 CC -!- PTM: Phosphorylated on serine residues in the C-terminal  
 CC cytoplasmic tail, essentially between residues 667 and 853.  
 CC -!- SIMILARITY: Belongs to the potassium channel family. B (Shab)  
 CC subfamily.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X16476; CAA34497.1; ALT\_INIT.  
 DR PIR; S05448; CHRTD1.  
 DR HSSP; Q54397; 1BL8.  
 DR InterPro; IPR000210; BTB\_POZ.  
 DR InterPro; IPR005821; Ion\_trans.  
 DR InterPro; IPR001622; K+channel\_pore.  
 DR InterPro; IPR003091; K\_channel.  
 DR InterPro; IPR003131; K\_tetra.  
 DR InterPro; IPR004350; Kv21channel.  
 DR InterPro; IPR003968; Kv\_channel.  
 DR InterPro; IPR005820; M+channel\_nlg.  
 DR InterPro; IPR003973; Shab\_channel.  
 DR Pfam; PF00520; ion\_trans; 1.  
 DR Pfam; PF02214; K\_tetra; 1.  
 DR Pfam; PF03521; Kv2channel; 1.  
 DR PRINTS; PR00169; KCHANNEL.  
 DR PRINTS; PR01514; KV21CHANNEL.  
 DR PRINTS; PR01491; KVCHANNEL.  
 DR PRINTS; PR01495; SHABCHANNEL.

DR SMART; SM00225; BTB; 1.  
KW Transport; Ion transport; Ionic channel; Voltage-gated channel;  
KW Potassium channel; Potassium; Potassium transport; Transmembrane;  
KW Phosphorylation; Multigene family.  
FT DOMAIN 1 186 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 187 208 SEGMENT S1 (POTENTIAL).  
FT TRANSMEM 229 250 SEGMENT S2 (POTENTIAL).  
FT DOMAIN 251 260 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 261 282 SEGMENT S3 (POTENTIAL).  
FT TRANSMEM 295 316 SEGMENT S4 (POTENTIAL).  
FT DOMAIN 317 330 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 331 352 SEGMENT S5 (POTENTIAL).  
FT DOMAIN 365 385 SEGMENT H5 (PORE-FORMING) (POTENTIAL).  
FT TRANSMEM 393 414 SEGMENT S6 (POTENTIAL).  
FT DOMAIN 415 857 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 517 520 POLY-SER.  
FT DOMAIN 700 705 POLY-ALA.  
FT MUTAGEN 444 444 S->A: NO EFFECT ON PHOSPHORYLATION.  
FT MUTAGEN 496 496 S->A: NO EFFECT ON PHOSPHORYLATION.  
SQ SEQUENCE 857 AA; 95637 MW; B3C5B0839AB15FD0 CRC64;

Query Match 70.7%; Score 29; DB 1; Length 857;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAE 6  
|||||  
Db 348 LVFFAE 353

# RESULT 53

## KCB1\_HUMAN

ID KCB1\_HUMAN STANDARD; PRT; 858 AA.  
AC Q14721; Q14193;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Potassium voltage-gated channel subfamily B member 1 (Potassium  
DE channel Kv2.1) (h-DRK1).  
GN KCNB1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94363205; PubMed=8081723;  
RA Albrecht B., Lorra C., Stocker K., Pongs O.;  
RT "Cloning and characterization of a human delayed rectifier potassium  
RT channel gene.";  
RL Recept. Channels 1:99-110(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Ikeda S.R., Soler F., Zuhlke R.D., Joho R.H., Lewis D.L.;  
RL Submitted (JAN-1993) to the EMBL/GenBank/DDBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.

RC TISSUE=Lens epithelium;  
 RA Rae J.L., Shepard A.R.;  
 RT "Identification of potassium channels in human lens epithelium.";  
 RL (In) Civan M.M. (eds.);  
 RL The eye's aqueous humor-from secretion to glaucoma, pp.69-104,  
 RL Academic Press, San Diego (1998).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21638749; PubMed=11780052;  
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Burrill W., Butler A.P., Carder C., Carter N.P.,  
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clegg S., Cobley V.E., Collier R.E., Connor R., Corby N.R.,  
 RA Coulson A., Coville G.J., Deadman R., Dhami P., Dunn M.,  
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Lehvaslaiho M.H., Liversha M., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,  
 RA Milne S., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,  
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.;  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871(2001).  
 CC -!- FUNCTION: Mediates the voltage-dependent potassium ion  
 CC permeability of excitable membranes. Channels open or close in  
 CC response to the voltage difference across the membrane, letting K+  
 CC ions pass in accordance with their electrochemical gradient.  
 CC -!- SUBUNIT: Heteromultimer with KCNG2, KCNG3, KCNG4, KCNS1, KCNS2,  
 CC KCNS3 and KCNV2 (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- DOMAIN: The segment S4 is probably the voltage-sensor and is  
 CC characterized by a series of positively charged amino acids at  
 CC every third position.  
 CC -!- DOMAIN: The tail may be important in modulation of channel  
 CC activity and/or targeting of the channel to specific subcellular  
 CC compartments.  
 CC -!- PTM: Phosphorylated on serine residues in the C-terminal  
 CC cytoplasmic tail (By similarity).  
 CC -!- SIMILARITY: Belongs to the potassium channel family. B (Shab)  
 CC subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X68302; CAA48374.1; -.  
DR EMBL; L02840; AAA36156.1; ALT\_INIT.  
DR EMBL; AF026005; AAB88808.1; -.  
DR EMBL; AL035685; CAB89417.1; -.  
DR PIR; S31761; S31761.  
DR HSSP; Q54397; 1BL8.  
DR Genew; HGNC:6231; KCNB1.  
DR MIM; 600397; -.  
DR InterPro; IPR000210; BTB\_POZ.  
DR InterPro; IPR005821; Ion\_trans.  
DR InterPro; IPR001622; K+channel\_pore.  
DR InterPro; IPR003091; K\_channel.  
DR InterPro; IPR003131; K\_tetra.  
DR InterPro; IPR004350; Kv21channel.  
DR InterPro; IPR003968; Kv\_channel.  
DR InterPro; IPR005820; M+channel\_nlg.  
DR InterPro; IPR003973; Shab\_channel.  
DR Pfam; PF00520; ion\_trans; 1.  
DR Pfam; PF02214; K\_tetra; 1.  
DR Pfam; PF03521; Kv2channel; 1.  
DR PRINTS; PR00169; KCHANNEL.  
DR PRINTS; PR01514; KV21CHANNEL.  
DR PRINTS; PR01491; KVCHANNEL.  
DR PRINTS; PR01495; SHABCHANNEL.  
DR SMART; SM00225; BTB; 1.  
KW Transport; Ion transport; Ionic channel; Voltage-gated channel;  
KW Potassium channel; Potassium; Potassium transport; Transmembrane;  
KW Phosphorylation; Multigene family.  
FT DOMAIN 1 186 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 187 208 SEGMENT S1 (POTENTIAL).  
FT TRANSMEM 229 250 SEGMENT S2 (POTENTIAL).  
FT DOMAIN 251 260 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 261 282 SEGMENT S3 (POTENTIAL).  
FT TRANSMEM 295 316 SEGMENT S4 (POTENTIAL).  
FT DOMAIN 317 330 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 331 352 SEGMENT S5 (POTENTIAL).  
FT DOMAIN 365 385 SEGMENT H5 (PORE-FORMING) (POTENTIAL).  
FT TRANSMEM 393 414 SEGMENT S6 (POTENTIAL).  
FT DOMAIN 415 858 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 517 520 POLY-SER.  
FT DOMAIN 701 706 POLY-ALA.  
SQ SEQUENCE 858 AA; 95877 MW; C4B426174ED0DEE4 CRC64;

Query Match 70.7%; Score 29; DB 1; Length 858;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAE 6  
|||  
Db 348 LVFFAE 353

RESULT 54

KCB1\_PIG

ID KCB1\_PIG STANDARD; PRT; 858 AA.  
AC 018868;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Potassium voltage-gated channel subfamily B member 1 (Potassium  
channel Kv2.1) (DRK1).  
GN KCNB1.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lens epithelium;  
RA Rae J.L., Shepard A.R.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.  
CC -!- FUNCTION: Mediates the voltage-dependent potassium ion  
CC permeability of excitable membranes. Channels open or close in  
CC response to the voltage difference across the membrane, letting K+  
CC ions pass in accordance with their electrochemical gradient (By  
CC similarity).  
CC -!- SUBUNIT: Heteromultimer with KCNG2, KCNG3, KCNG4, KCNS1, KCNS2,  
CC KCNS3 and KCNV2 (By similarity).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- DOMAIN: The segment S4 is probably the voltage-sensor and is  
CC characterized by a series of positively charged amino acids at  
CC every third position.  
CC -!- DOMAIN: The tail may be important in modulation of channel  
CC activity and/or targeting of the channel to specific subcellular  
CC compartments.  
CC -!- PTM: Phosphorylated on serine residues in the C-terminal  
CC cytoplasmic tail (By similarity).  
CC -!- SIMILARITY: Belongs to the potassium channel family. B (Shab)  
CC subfamily.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF026006; AAB88809.1; -.  
DR HSSP; Q54397; 1BL8.  
DR InterPro; IPR000210; BTB\_POZ.  
DR InterPro; IPR005821; Ion\_trans.  
DR InterPro; IPR001622; K+channel\_pore.  
DR InterPro; IPR003091; K\_channel.  
DR InterPro; IPR003131; K\_tetra.  
DR InterPro; IPR004350; Kv2lchannel.  
DR InterPro; IPR003968; Kv\_channel.  
DR InterPro; IPR005820; M+channel\_nlg.  
DR InterPro; IPR003973; Shab\_channel.

DR Pfam; PF00520; ion\_trans; 1.  
 DR Pfam; PF02214; K\_tetra; 1.  
 DR Pfam; PF03521; Kv2channel; 1.  
 DR PRINTS; PR00169; KCHANNEL.  
 DR PRINTS; PR01514; KV21CHANNEL.  
 DR PRINTS; PR01491; KVCHANNEL.  
 DR PRINTS; PR01495; SHABCHANNEL.  
 DR SMART; SM00225; BTB; 1.  
 KW Transport; Ion transport; Ionic channel; Voltage-gated channel;  
 KW Potassium channel; Potassium; Potassium transport; Transmembrane;  
 KW Phosphorylation; Multigene family.  
 FT DOMAIN 1 185 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 186 206 SEGMENT S1 (POTENTIAL).  
 FT TRANSMEM 229 249 SEGMENT S2 (POTENTIAL).  
 FT DOMAIN 250 259 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 260 280 SEGMENT S3 (POTENTIAL).  
 FT TRANSMEM 295 315 SEGMENT S4 (POTENTIAL).  
 FT DOMAIN 316 330 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 331 351 SEGMENT S5 (POTENTIAL).  
 FT DOMAIN 365 385 SEGMENT H5 (PORE-FORMING) (POTENTIAL).  
 FT TRANSMEM 393 413 SEGMENT S6 (POTENTIAL).  
 FT DOMAIN 414 858 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 517 520 POLY-SER.  
 FT DOMAIN 701 706 POLY-ALA.  
 SQ SEQUENCE 858 AA; 96117 MW; A9E24C3A8E13B491 CRC64;  
  
 Query Match 70.7%; Score 29; DB 1; Length 858;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 Qy 1 LVFFAE 6  
 |||||  
 Db 348 LVFFAE 353

RESULT 55  
 KCB1\_RABIT  
 ID KCB1\_RABIT STANDARD; PRT; 858 AA.  
 AC Q9MZ19;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Potassium voltage-gated channel subfamily B member 1 (Potassium  
 DE channel Kv2.1).  
 GN KCNB1.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=New Zealand white; TISSUE=Corneal endothelium;  
 RA Rae J.L.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: Mediates the voltage-dependent potassium ion  
 CC permeability of excitable membranes. Channels open or close in  
 CC response to the voltage difference across the membrane, letting K+

ions pass in accordance with their electrochemical gradient (By similarity).  
 -!- SUBUNIT: Heteromultimer with KCNG2, KCNG3, KCNG4, KCNS1, KCNS2, KCNS3 and KCNV2 (By similarity).  
 -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 -!- DOMAIN: The segment S4 is probably the voltage-sensor and is characterized by a series of positively charged amino acids at every third position.  
 -!- DOMAIN: The tail may be important in modulation of channel activity and/or targeting of the channel to specific subcellular compartments.  
 -!- PTM: Phosphorylated on serine residues in the C-terminal cytoplasmic tail (By similarity).  
 -!- SIMILARITY: Belongs to the potassium channel family. B (Shab) subfamily.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

DR EMBL; AF266507; AAF77058.1; -.  
 DR HSSP; Q54397; 1BL8.  
 DR InterPro; IPR000210; BTB\_POZ.  
 DR InterPro; IPR005821; Ion\_trans.  
 DR InterPro; IPR001622; K+channel\_pore.  
 DR InterPro; IPR003091; K\_channel.  
 DR InterPro; IPR003131; K\_tetra.  
 DR InterPro; IPR004350; Kv21channel.  
 DR InterPro; IPR003968; Kv\_channel.  
 DR InterPro; IPR005820; M+channel\_nlg.  
 DR InterPro; IPR003973; Shab\_channel.  
 DR Pfam; PF00520; ion\_trans; 1.  
 DR Pfam; PF02214; K\_tetra; 1.  
 DR Pfam; PF03521; Kv2channel; 1.  
 DR PRINTS; PR00169; KCHANNEL.  
 DR PRINTS; PR01514; KV21CHANNEL.  
 DR PRINTS; PR01491; KVCHANNEL.  
 DR PRINTS; PR01495; SHABCHANNEL.  
 DR SMART; SM00225; BTB; 1.  
 KW Transport; Ion transport; Ionic channel; Voltage-gated channel;  
 KW Potassium channel; Potassium; Potassium transport; Transmembrane;  
 KW Phosphorylation; Multigene family.  
 FT DOMAIN 1 185 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 186 206 SEGMENT S1 (POTENTIAL).  
 FT TRANSMEM 229 249 SEGMENT S2 (POTENTIAL).  
 FT DOMAIN 250 259 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 260 280 SEGMENT S3 (POTENTIAL).  
 FT TRANSMEM 295 315 SEGMENT S4 (POTENTIAL).  
 FT DOMAIN 316 330 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 331 351 SEGMENT S5 (POTENTIAL).  
 FT DOMAIN 365 385 SEGMENT H5 (PORE-FORMING) (POTENTIAL).  
 FT TRANSMEM 393 413 SEGMENT S6 (POTENTIAL).  
 FT DOMAIN 414 858 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 517 520 POLY-SER.  
 FT DOMAIN 701 706 POLY-ALA.  
 SQ SEQUENCE 858 AA; 95373 MW; FDFDCC49B6973C4B CRC64;

Query Match 70.7%; Score 29; DB 1; Length 858;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVFFAE 6  
 |||||  
 Db 348 LVFFAE 353

# RESULT 56

## KCB2\_RAT

ID KCB2\_RAT STANDARD; PRT; 907 AA.  
 AC Q63099;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Potassium voltage-gated channel subfamily B member 2 (Potassium channel Kv2.2) (CDRK).  
 GN KCNB2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=92198655; PubMed=1550672;  
 RA Hwang P.M., Glatt C.E., Bredt D.S., Yellen G., Snyder S.H.;  
 RT "A novel K+ channel with unique localizations in mammalian brain:  
 RT molecular cloning and characterization."  
 RL Neuron 8:473-481(1992).  
 RN [2]  
 RP CONCEPTUAL TRANSLATION.  
 RA Sundaram S.;  
 RL Unpublished observations (OCT-2002).  
 CC -!- FUNCTION: Mediates the voltage-dependent potassium ion  
 CC permeability of excitable membranes. Channels open or close in  
 CC response to the voltage difference across the membrane, letting K+  
 CC ions pass in accordance with their electrochemical gradient.  
 CC -!- SUBUNIT: Heteromultimer with KCNS1, KCNS2 and KCNS3 (By  
 CC similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- TISSUE SPECIFICITY: In the brain, the greatest density occurs in  
 CC the olfactory bulb, followed by the cerebral cortex, hippocampus,  
 CC and cerebellum. In peripheral tissues it is most prominent in  
 CC whole tongue epithelium and circumvallate papillae.  
 CC -!- DOMAIN: The segment S4 is probably the voltage-sensor and is  
 CC characterized by a series of positively charged amino acids at  
 CC every third position.  
 CC -!- DOMAIN: The tail may be important in modulation of channel  
 CC activity and/or targeting of the channel to specific subcellular  
 CC compartments.  
 CC -!- PTM: Phosphorylated (Potential).



```

CC  -!- SIMILARITY: Belongs to the potassium channel family. B (Shab)
CC      subfamily.
CC  -!- CAUTION: This is a conceptual translation, a frameshift was
CC      introduced in position 755 to restore the similarity with the
CC      other orthologs.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; M77482; AAA40905.1; ALT_FRAME.
DR  HSSP; Q54397; 1BL8.
DR  InterPro; IPR000210; BTB_POZ.
DR  InterPro; IPR005821; Ion_trans.
DR  InterPro; IPR001622; K+channel_pore.
DR  InterPro; IPR003091; K_channel.
DR  InterPro; IPR003131; K_tetra.
DR  InterPro; IPR005826; Kv22channel.
DR  InterPro; IPR003968; Kv_channel.
DR  InterPro; IPR005820; M+channel_nlg.
DR  InterPro; IPR003973; Shab_channel.
DR  Pfam; PF00520; ion_trans; 1.
DR  Pfam; PF02214; K_tetra; 1.
DR  Pfam; PF03521; Kv2channel; 1.
DR  PRINTS; PR00169; KCHANNEL.
DR  PRINTS; PR01515; KV22CHANNEL.
DR  PRINTS; PR01491; KVCHANNEL.
DR  PRINTS; PR01495; SHABCHANNEL.
DR  SMART; SM00225; BTB; 1.
KW  Transport; Ion transport; Ionic channel; Voltage-gated channel;
KW  Potassium channel; Potassium; Potassium transport; Transmembrane;
KW  Phosphorylation; Glycoprotein; Multigene family.
FT  DOMAIN      1      190      CYTOPLASMIC (POTENTIAL).
FT  TRANSMEM    191    212      SEGMENT S1 (POTENTIAL).
FT  TRANSMEM    233    254      SEGMENT S2 (POTENTIAL).
FT  DOMAIN      255    264      CYTOPLASMIC (POTENTIAL).
FT  TRANSMEM    265    286      SEGMENT S3 (POTENTIAL).
FT  TRANSMEM    299    320      SEGMENT S4 (POTENTIAL).
FT  DOMAIN      321    334      CYTOPLASMIC (POTENTIAL).
FT  TRANSMEM    335    356      SEGMENT S5 (POTENTIAL).
FT  DOMAIN      369    389      SEGMENT H5 (PORE-FORMING) (POTENTIAL).
FT  TRANSMEM    397    418      SEGMENT S6 (POTENTIAL).
FT  DOMAIN      419    907      CYTOPLASMIC (POTENTIAL).
FT  CARBOHYD    287    287      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ  SEQUENCE    907 AA;  102095 MW;  B242D9A6753A1295 CRC64;

```

```

Query Match          70.7%; Score 29; DB 1; Length 907;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches      6; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 LVFFAE 6
        |||||
Db     352 LVFFAE 357

```

RESULT 57

KCB2\_HUMAN

ID KCB2\_HUMAN STANDARD; PRT; 911 AA.  
AC Q92953; Q9BXD3;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Potassium voltage-gated channel subfamily B member 2 (Potassium  
DE channel Kv2.2).  
GN KCNB2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE OF 1-811 FROM N.A.  
RX MEDLINE=98275219; PubMed=9612272;  
RA Schmalz F., Kinsella J., Koh S.D., Vogalis F., Schneider A.,  
RA Flynn E.R., Kenyon J.L., Horowitz B.;  
RT "Molecular identification of a component of delayed rectifier current  
RT in gastrointestinal smooth muscles.";  
RL Am. J. Physiol. 274:G901-G911(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Rae J.L.;  
RL Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.  
CC -!- FUNCTION: Mediates the voltage-dependent potassium ion  
CC permeability of excitable membranes. Channels open or close in  
CC response to the voltage difference across the membrane, letting K+  
CC ions pass in accordance with their electrochemical gradient.  
CC -!- SUBUNIT: Heteromultimer with KCNS1, KCNS2 and KCNS3 (By  
CC similarity).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- DOMAIN: The segment S4 is probably the voltage-sensor and is  
CC characterized by a series of positively charged amino acids at  
CC every third position.  
CC -!- DOMAIN: The tail may be important in modulation of channel  
CC activity and/or targeting of the channel to specific subcellular  
CC compartments.  
CC -!- PTM: Phosphorylated (Potential).  
CC -!- SIMILARITY: Belongs to the potassium channel family. B (Shab)  
CC subfamily.  
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a  
CC frameshift in position 768.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U69962; AAB08433.1; ALT\_FRAME.  
DR EMBL; AF338730; AAK16585.1; -.

DR HSSP; Q54397; 1BL8.  
 DR Genew; HGNC:6232; KCNB2.  
 DR MIM; 607738; -.  
 DR GO; GO:0008076; C:voltage-gated potassium channel complex; TAS.  
 DR GO; GO:0005251; F:delayed rectifier potassium channel activity; TAS.  
 DR GO; GO:0006813; P:potassium ion transport; TAS.  
 DR GO; GO:0006940; P:regulation of smooth muscle contraction; TAS.  
 DR InterPro; IPR000210; BTB\_POZ.  
 DR InterPro; IPR005821; Ion\_trans.  
 DR InterPro; IPR001622; K+channel\_pore.  
 DR InterPro; IPR003091; K\_channel.  
 DR InterPro; IPR003131; K\_tetra.  
 DR InterPro; IPR005826; Kv22channel.  
 DR InterPro; IPR003968; Kv\_channel.  
 DR InterPro; IPR005820; M+channel\_nlg.  
 DR InterPro; IPR003973; Shab\_channel.  
 DR Pfam; PF00520; ion\_trans; 1.  
 DR Pfam; PF02214; K\_tetra; 1.  
 DR Pfam; PF03521; Kv2channel; 1.  
 DR PRINTS; PR00169; KCHANNEL.  
 DR PRINTS; PR01515; KV22CHANNEL.  
 DR PRINTS; PR01491; KVCHANNEL.  
 DR PRINTS; PR01495; SHABCHANNEL.  
 DR SMART; SM00225; BTB; 1.  
 KW Transport; Ion transport; Ionic channel; Voltage-gated channel;  
 KW Potassium channel; Potassium; Potassium transport; Transmembrane;  
 KW Phosphorylation; Glycoprotein; Multigene family.  
 FT DOMAIN 1 190 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 191 212 SEGMENT S1 (POTENTIAL).  
 FT TRANSMEM 233 254 SEGMENT S2 (POTENTIAL).  
 FT DOMAIN 255 264 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 265 286 SEGMENT S3 (POTENTIAL).  
 FT TRANSMEM 299 320 SEGMENT S4 (POTENTIAL).  
 FT DOMAIN 321 334 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 335 356 SEGMENT S5 (POTENTIAL).  
 FT DOMAIN 369 389 SEGMENT H5 (PORE-FORMING) (POTENTIAL).  
 FT TRANSMEM 397 418 SEGMENT S6 (POTENTIAL).  
 FT DOMAIN 419 911 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 911 AA; 102562 MW; 2261D7D6280CF81A CRC64;

Query Match 70.7%; Score 29; DB 1; Length 911;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVFFAE 6  
 |||||  
 Db 352 LVFFAE 357

# RESULT 58

KCB2\_RABIT

ID KCB2\_RABIT STANDARD; PRT; 911 AA.  
 AC Q95L11;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Potassium voltage-gated channel subfamily B member 2 (Potassium  
 DE channel Kv2.2).  
 GN KCNB2.  
 OS *Oryctolagus cuniculus* (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; *Oryctolagus*.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22286512; PubMed=12399537;  
 RA Malysz J., Farrugia G., Ou Y., Szurszewski J.H., Nehra A.,  
 RA Gibbons S.J.;  
 RT "The Kv2.2 alpha subunit contributes to delayed rectifier K(+) currents in myocytes from rabbit corpus cavernosum.";  
 RT J. Androl. 23:899-910(2002).  
 RL J. Androl. 23:899-910(2002).  
 CC -!- FUNCTION: Mediates the voltage-dependent potassium ion permeability of excitable membranes. Channels open or close in response to the voltage difference across the membrane, letting K+ ions pass in accordance with their electrochemical gradient (By similarity).  
 CC -!- SUBUNIT: Heteromultimer with KCNS1, KCNS2 and KCNS3 (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- DOMAIN: The segment S4 is probably the voltage-sensor and is characterized by a series of positively charged amino acids at every third position.  
 CC -!- DOMAIN: The tail may be important in modulation of channel activity and/or targeting of the channel to specific subcellular compartments.  
 CC -!- PTM: Phosphorylated (Potential).  
 CC -!- SIMILARITY: Belongs to the potassium channel family. B (Shab) subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AY037947; AAK84954.1; -.  
 DR InterPro; IPR000210; BTB\_POZ.  
 DR InterPro; IPR005821; Ion\_trans.  
 DR InterPro; IPR001622; K+channel\_pore.  
 DR InterPro; IPR003091; K\_channel.  
 DR InterPro; IPR003131; K\_tetra.  
 DR InterPro; IPR005826; Kv22channel.  
 DR InterPro; IPR003968; Kv\_channel.  
 DR InterPro; IPR005820; M+channel\_nlg.  
 DR InterPro; IPR003973; Shab\_channel.  
 DR Pfam; PF00520; ion\_trans; 1.  
 DR Pfam; PF02214; K\_tetra; 1.  
 DR Pfam; PF03521; Kv2channel; 1.  
 DR PRINTS; PR00169; KCHANNEL.  
 DR PRINTS; PR01515; KV22CHANNEL.  
 DR PRINTS; PR01491; KVCHANNEL.

DR PRINTS; PR01495; SHABCHANNEL.  
 DR SMART; SM00225; BTB; 1.  
 KW Transport; Ion transport; Ionic channel; Voltage-gated channel;  
 KW Potassium channel; Potassium; Potassium transport; Transmembrane;  
 KW Phosphorylation; Glycoprotein; Multigene family.  
 FT DOMAIN 1 189 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 190 210 SEGMENT S1 (POTENTIAL).  
 FT TRANSMEM 233 253 SEGMENT S2 (POTENTIAL).  
 FT DOMAIN 254 263 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 264 284 SEGMENT S3 (POTENTIAL).  
 FT TRANSMEM 299 320 SEGMENT S4 (POTENTIAL).  
 FT DOMAIN 321 334 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 335 356 SEGMENT S5 (POTENTIAL).  
 FT DOMAIN 369 389 SEGMENT H5 (PORE-FORMING) (POTENTIAL).  
 FT TRANSMEM 397 418 SEGMENT S6 (POTENTIAL).  
 FT DOMAIN 419 911 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 911 AA; 102278 MW; 69353D0664C5D689 CRC64;

Query Match 70.7%; Score 29; DB 1; Length 911;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAE 6  
 |||||  
 Db 352 LVFFAE 357

# RESULT 59

## SYI\_ARCFU

ID SYI\_ARCFU STANDARD; PRT; 1018 AA.  
 AC O29622;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Isoleucyl-tRNA synthetase (EC 6.1.1.5) (Isoleucine--tRNA ligase)  
 DE (IleRS).  
 GN ILES OR AF0633.  
 OS Archaeoglobus fulgidus.  
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
 OC Archaeoglobaceae; Archaeoglobus.  
 OX NCBI\_TaxID=2234;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE=98049343; PubMed=9389475;  
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,  
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
 RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;

RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon *Archaeoglobus fulgidus*.";  
 RL Nature 390:364-370(1997).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-isoleucine + tRNA(Ile) = AMP +  
 CC diphosphate + L-isoleucyl-tRNA(Ile).  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE001061; AAB90608.1; -.  
 DR PIR; A69329; A69329.  
 DR HSSP; P56690; 1ILE.  
 DR TIGR; AF0633; -.  
 DR InterPro; IPR002300; tRNA-synt\_1a.  
 DR InterPro; IPR001412; tRNA-synt\_I.  
 DR InterPro; IPR002301; tRNA-synt\_ile.  
 DR InterPro; IPR009008; ValRS\_IleRS\_edit.  
 DR Pfam; PF00133; tRNA-synt\_1; 1.  
 DR PRINTS; PR00984; TRNASYNTHILE.  
 DR TIGRFAMS; TIGR00392; ileS; 1.  
 DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; 1.  
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 KW Metal-binding; Zinc; Complete proteome.  
 FT SITE 43 53 "HIGH" REGION.  
 FT SITE 586 590 "KMSKS" REGION.  
 FT BINDING 589 589 ATP (BY SIMILARITY).  
 SQ SEQUENCE 1018 AA; 118884 MW; 58C8BBE472FF1D0F CRC64;

Query Match 70.7%; Score 29; DB 1; Length 1018;  
 Best Local Similarity 83.3%; Pred. No. 2.6e+02;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 FFAEDF 8  
 || |||  
 Db 700 FFVEDF 705

# RESULT 60

## VNS2\_CVBM

ID VNS2\_CVBM STANDARD; PRT; 109 AA.  
 AC P15774;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Nonstructural protein 2 (Nonstructural 12.7 kDa protein).  
 GN NS2.  
 OS Bovine coronavirus (strain Mebus) (BCoV) (BCV), and  
 OS Bovine coronavirus (strain F15) (BCoV) (BCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;

OC Coronaviridae; Coronavirus.  
 OX NCBI\_TaxID=11132, 11129;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Mebus;  
 RX MEDLINE=90320120; PubMed=2142556;  
 RA Abraham S., Kienzle T.E., Lapps W.E., Brian D.A.;  
 RT "Sequence and expression analysis of potential nonstructural proteins  
 RT of 4.9, 4.8, 12.7, and 9.5 kDa encoded between the spike and membrane  
 RT protein genes of the bovine coronavirus.";  
 RL Virology 177:488-495(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=F15;  
 RX MEDLINE=90206809; PubMed=2320429;  
 RA Woloszyn N., Boireau P., Laporte J.;  
 RT "Nucleotide sequence of the bovine enteric coronavirus BECV F15 mRNA  
 RT 5 and mRNA 6 unique regions.";  
 RL Nucleic Acids Res. 18:1303-1303(1990).  
 CC -!- SIMILARITY: Belongs to the coronavirus NS2 protein family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X51347; CAA35740.1; -.  
 DR EMBL; U00735; AAA42913.1; -.  
 DR PIR; S08408; MNIHB2.  
 DR InterPro; IPR006841; Corona\_NS2.  
 DR Pfam; PF04753; Corona\_NS2; 1.  
 KW Nonstructural protein.  
 SQ SEQUENCE 109 AA; 12806 MW; 08B7CA339A1BD051 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 109;  
 Best Local Similarity 75.0%; Pred. No. 46;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 ||| | ||  
 Db 90 LVFLAVDF 97

# RESULT 61

## VNS2\_CVHOC

ID VNS2\_CVHOC STANDARD; PRT; 109 AA.  
 AC Q04853;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Nonstructural protein 2 (Nonstructural 12.9 kDa protein).  
 GN NS2.  
 OS Human coronavirus (strain OC43) (HCoV-OC43).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;

OC Coronaviridae; Coronavirus.  
 OX NCBI\_TaxID=31631;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93297129; PubMed=8517026;  
 RA Mounir S., Talbot P.J.;  
 RT "Human coronavirus OC43 RNA 4 lacks two open reading frames located  
 RT downstream of the S gene of bovine coronavirus.";  
 RL Virology 192:355-360(1993).  
 CC -!- FUNCTION: May act as membrane-anchoring region for structural  
 CC proteins during virus assembly, or play a role in membrane  
 CC association of the viral polymerase during replication.  
 CC -!- SIMILARITY: Belongs to the coronavirus NS2 protein family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M99576; AAA02569.1; -.  
 DR PIR; A44275; A44275.  
 DR InterPro; IPR006841; Corona\_NS2.  
 DR Pfam; PF04753; Corona\_NS2; 1.  
 KW Nonstructural protein.  
 SQ SEQUENCE 109 AA; 12935 MW; 8F06A6E9E30B1667 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 109;  
 Best Local Similarity 75.0%; Pred. No. 46;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 LVFFAEDF 8  
 ||| | ||  
 Db 90 LVFLAVDF 97

# RESULT 62

## RL7A\_HALMA

ID RL7A HALMA STANDARD; PRT; 119 AA.  
 AC P12743;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE 50S ribosomal protein L7Ae (HS6).  
 GN RPL7AE.  
 OS Haloarcula marismortui (Halobacterium marismortui).  
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 OC Halobacteriaceae; Haloarcula.  
 OX NCBI\_TaxID=2238;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=88055606; PubMed=3315748;  
 RA Kimura J., Arndt E., Kimura M.;  
 RT "Primary structures of three highly acidic ribosomal proteins S6, S12  
 RT and S15 from the archaebacterium Halobacterium marismortui.";



RL FEBS Lett. 224:65-70(1987).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS), AND REVISIONS TO 108-110.  
 RC STRAIN=ATCC 43049;  
 RX MEDLINE=20396344; PubMed=10937989;  
 RA Ban N., Nissen P., Hansen J., Moore P.B., Steitz T.A.;  
 RT "The complete atomic structure of the large ribosomal subunit at 2.4  
 RT A resolution."  
 RL Science 289:905-920(2000).  
 CC -!- SIMILARITY: Belongs to the L7AE family of ribosomal proteins.  
 CC -!- CAUTION: Was originally (Ref.1) thought to be a protein from the  
 CC small subunit of the ribosome.  
 DR PIR; S00182; R5HSS6.  
 DR PDB; 1FFK; 26-SEP-01.  
 DR PDB; 1K8A; 19-JUL-02.  
 DR PDB; 1K9M; 19-JUL-02.  
 DR PDB; 1KD1; 19-JUL-02.  
 DR PDB; 1M1K; 23-AUG-02.  
 DR PDB; 1M90; 06-SEP-02.  
 DR HAMAP; MF\_00326; -; 1.  
 DR InterPro; IPR004038; Ribosomal\_L7A.  
 DR InterPro; IPR004037; Ribosomal\_L7Ae.  
 DR Pfam; PF01248; Ribosomal\_L7Ae; 1.  
 DR PRINTS; PR00881; L7ARS6FAMILY.  
 DR PROSITE; PS01082; RIBOSOMAL\_L7AE; 1.  
 KW Ribosomal protein; 3D-structure.  
 FT INIT\_MET 0 0  
 FT CONFLICT 108 110 MISSING (IN REF. 1).  
 SQ SEQUENCE 119 AA; 12591 MW; A132AA40519F45C5 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 119;  
 Best Local Similarity 85.7%; Pred. No. 50;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LVFFAED 7  
 ||| |||  
 Db 47 LVFVAED 53

# RESULT 63

Y328\_SYNY3

ID Y328\_SYNY3 STANDARD; PRT; 157 AA.  
 AC Q55535;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Putative low molecular weight protein-tyrosine-phosphatase slr0328  
 DE (EC 3.1.3.48).  
 GN SLR0328.  
 OS Synechocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 OX NCBI\_TaxID=1148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96127529; PubMed=8590279;  
 RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,  
 RA Sugiura M., Tabata S.;

RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb  
 RT region from map positions 64% to 92% of the genome.";  
 RL DNA Res. 2:153-166(1995).  
 CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein  
 CC tyrosine + phosphate.  
 CC -!- SIMILARITY: Belongs to the low molecular weight phosphotyrosine  
 CC protein phosphatase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D63999; BAA10030.1; -.  
 DR PIR; S76052; S76052.  
 DR HSSP; P11064; 1DG9.  
 DR InterPro; IPR000106; Low\_mwt\_PTPase.  
 DR Pfam; PF01451; LMWPC; 1.  
 DR PRINTS; PR00719; LMWPTPASE.  
 DR SMART; SM00226; LMWPC; 1.  
 KW Hypothetical protein; Hydrolase; Complete proteome.  
 FT ACT\_SITE 7 7 NUCLEOPHILE (BY SIMILARITY).  
 FT ACT\_SITE 12 12 BY SIMILARITY.  
 FT ACT\_SITE 124 124 PROTON DONOR (BY SIMILARITY).  
 SQ SEQUENCE 157 AA; 17470 MW; D642D8E6D3A49751 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 157;  
 Best Local Similarity 83.3%; Pred. No. 66;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 FFAEDF 8  
 || |||  
 Db 72 FFPEDF 77

# RESULT 64

## DESS\_MYXXA

ID DESS\_MYXXA STANDARD; PRT; 173 AA.  
 AC P02966;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Development-specific protein S (Spore coat protein S).  
 GN TPS.  
 OS Myxococcus xanthus.  
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;  
 OC Cystobacterineae; Myxococcaceae; Myxococcus.  
 OX NCBI\_TaxID=34;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84070723; PubMed=6316328;  
 RA Inouye S., Franceschini T., Inouye M.;  
 RT "Structural similarities between the development-specific protein S

RT from a Gram-negative bacterium, *Myxococcus xanthus*, and calmodulin.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:6829-6833(1983).  
 RN [2]  
 RP CALCIUM-BINDING SITES, AND MUTAGENESIS.  
 RX MEDLINE=88087248; PubMed=3121626;  
 RA Teintze M., Inouye M., Inouye S.;  
 RT "Characterization of calcium-binding sites in development-specific  
 RT protein S of *Myxococcus xanthus* using site-specific mutagenesis.";  
 RL J. Biol. Chem. 263:1199-1203(1988).  
 RN [3]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE=94240126; PubMed=8183906;  
 RA Bagby S., Harvey T.S., Eagle S.G., Inouye S., Ikura M.;  
 RT "Structural similarity of a developmentally regulated bacterial spore  
 RT coat protein to beta gamma-crystallins of the vertebrate eye lens.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:4308-4312(1994).  
 RN [4]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE=94363227; PubMed=8081742;  
 RA Bagby S., Harvey T.S., Eagle S.G., Inouye S., Ikura M.;  
 RT "NMR-derived three-dimensional solution structure of protein S  
 RT complexed with calcium.";  
 RL Structure 2:107-122(1994).  
 CC -!- FUNCTION: PROTEIN S, INDUCED IN LARGE AMOUNTS DURING FRUITING BODY  
 CC FORMATION, ASSEMBLES ON THE SURFACE OF MYXOSPORES IN THE PRESENCE  
 CC OF CALCIUM ION. IT BINDS TWO MOLES OF CALCIUM.  
 CC -!- DOMAIN: Has a two-domain beta-structure, folded into four very  
 CC similar greek key motifs.  
 CC -!- SIMILARITY: Belongs to the beta/gamma-crystallin family.  
 CC -!- SIMILARITY: Contains 4 beta/gamma crystallin 'Greek key' domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; J01745; AAA25407.1; -.  
 DR PIR; A03490; DZYZSX.  
 DR PDB; 1PRR; 31-AUG-94.  
 DR PDB; 1PRS; 31-AUG-94.  
 DR PDB; 1NPS; 10-JAN-01.  
 DR InterPro; IPR001064; Crystallin.  
 DR Pfam; PF00030; crystall; 2.  
 DR SMART; SM00247; XTALbg; 2.  
 DR PROSITE; PS50915; CRYSTALLIN\_BETAGAMMA; 4.  
 KW Calcium-binding; Repeat; Sporulation; 3D-structure.  
 FT DOMAIN 2 46 BETA/GAMMA CRYSTALLIN 'GREEK KEY' 1.  
 FT DOMAIN 48 86 BETA/GAMMA CRYSTALLIN 'GREEK KEY' 2.  
 FT DOMAIN 87 90 CONNECTING PEPTIDE.  
 FT DOMAIN 91 135 BETA/GAMMA CRYSTALLIN 'GREEK KEY' 3.  
 FT DOMAIN 136 173 BETA/GAMMA CRYSTALLIN 'GREEK KEY' 4.  
 FT CA\_BIND 32 46 PROBABLE.  
 FT CA\_BIND 121 135 PROBABLE.  
 FT MUTAGEN 40 40 S->R: SUPPRESSES COMPLETELY CA-BINDING.

FT	MUTAGEN	129	129	S->R: LOWER CALCIUM AFFINITY.
FT	STRAND	4	7	
FT	TURN	11	12	
FT	STRAND	16	19	
FT	STRAND	22	24	
FT	HELIX	26	32	
FT	TURN	37	38	
FT	STRAND	41	43	
FT	STRAND	48	53	
FT	TURN	56	58	
FT	STRAND	62	65	
FT	STRAND	68	69	
FT	TURN	70	71	
FT	HELIX	74	77	
FT	STRAND	81	86	
FT	STRAND	93	95	
FT	TURN	99	101	
FT	STRAND	106	108	
FT	STRAND	111	113	
FT	TURN	115	118	
FT	HELIX	119	122	
FT	STRAND	131	132	
FT	STRAND	138	141	
FT	HELIX	144	146	
FT	STRAND	150	153	
FT	STRAND	157	157	
FT	TURN	161	163	
FT	STRAND	169	173	
SQ	SEQUENCE	173 AA; 18793 MW; C504A6409752C454 CRC64;		

Query Match 68.3%; Score 28; DB 1; Length 173;  
 Best Local Similarity 71.4%; Pred. No. 73;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy	2	VFFAEDF	8
		:	
Db	6	VFYNEDE	12

# RESULT 65

DEST\_MYXXA

ID DEST\_MYXXA STANDARD; PRT; 175 AA.

AC P02967;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Development-specific protein S homolog.

GN OPS.

OS Myxococcus xanthus.

OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;

OC Cystobacterineae; Myxococcaceae; Myxococcus.

OX NCBI\_TaxID=34;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=84070723; PubMed=6316328;

RA Inouye S., Franceschini T., Inouye M.;

RT "Structural similarities between the development-specific protein S

RT from a Gram-negative bacterium, *Myxococcus xanthus*, and calmodulin.";

RL Proc. Natl. Acad. Sci. U.S.A. 80:6829-6833(1983).

CC -!- SUBCELLULAR LOCATION: INSIDE THE SPORES.

CC -!- DOMAIN: Has a two-domain beta-structure, folded into four very similar greek key motifs.

CC -!- SIMILARITY: Belongs to the beta/gamma-crystallin family.

CC -!- SIMILARITY: Contains 4 beta/gamma crystallin 'Greek key' domains.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; J01745; AAA25406.1; -.

DR PIR; A03491; DZYZ1X.

DR HSSP; P02966; 1PRS.

DR InterPro; IPR001064; Crystallin.

DR Pfam; PF00030; crystall; 2.

DR SMART; SM00247; XTALbg; 2.

DR PROSITE; PS50915; CRYSTALLIN\_BETAGAMMA; 4.

KW Sporulation; Repeat.

FT	DOMAIN	2	46	BETA/GAMMA CRYSTALLIN 'GREEK KEY' 1.
FT	DOMAIN	48	86	BETA/GAMMA CRYSTALLIN 'GREEK KEY' 2.
FT	DOMAIN	87	90	CONNECTING PEPTIDE.
FT	DOMAIN	91	135	BETA/GAMMA CRYSTALLIN 'GREEK KEY' 3.
FT	DOMAIN	136	175	BETA/GAMMA CRYSTALLIN 'GREEK KEY' 4.

SQ SEQUENCE 175 AA; 19235 MW; 9514FE18D373F9B7 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 175;

Best Local Similarity 71.4%; Pred. No. 73;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
 ||: |||  
 Db 6 VFYNEDF 12

# RESULT 66

CME2\_BACSU

ID CME2\_BACSU STANDARD; PRT; 189 AA.

AC P32393;

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE ComE operon protein 2.

GN COMEB OR COME2 OR BSU25580.

OS *Bacillus subtilis*.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; *Bacillus*.

OX NCBI\_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95058187; PubMed=7968523;

RA Hahn J., Inamine G., Kozlov Y., Dubnau D.A.;

RT "Characterization of comE, a late competence operon of *Bacillus*

RT subtilis required for the binding and uptake of transforming DNA.";  
 RL Mol. Microbiol. 10:99-110(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / JH642;  
 RX MEDLINE=97124195; PubMed=8969508;  
 RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,  
 RA Kobayashi Y.;  
 RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of  
 RT the Bacillus subtilis genome containing the skin element and many  
 RT sporulation genes.";  
 RL Microbiology 142:3103-3111(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 RT subtilis.";  
 RL Nature 390:249-256(1997).  
 CC -!- FUNCTION: DISPENSABLE FOR TRANSFORMABILITY.  
 CC -!- COFACTOR: Zinc (By similarity).  
 CC -!- SIMILARITY: Belongs to the cytidine and deoxycytidylate deaminases  
 CC family.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; L15202; AAC36906.1; -.  
DR EMBL; D84432; BAA12453.1; -.  
DR EMBL; Z99117; CAB14500.1; -.  
DR PIR; S39864; S39864.  
DR SubtiList; BG10481; comEB.  
DR InterPro; IPR002125; dCMP\_cyt\_deam.  
DR Pfam; PF00383; dCMP\_cyt\_deam; 1.  
DR PROSITE; PS00903; CYT\_DCMF\_DEAMINASES; 1.  
KW Hydrolase; Zinc; Complete proteome.  
FT METAL 70 70 ZINC (BY SIMILARITY).  
FT METAL 98 98 ZINC (BY SIMILARITY).  
FT METAL 101 101 ZINC (BY SIMILARITY).  
SQ SEQUENCE 189 AA; 20969 MW; B6AC951966EE2D37 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 189;  
Best Local Similarity 57.1%; Pred. No. 79;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
|::|||:  
Db 113 VYYAEDY 119

#### RESULT 67

##### BCB1\_ARATH

ID BCB1\_ARATH STANDARD; PRT; 196 AA.  
AC Q07488; O82664;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Blue copper protein precursor (Blue copper-binding protein) (AtBCB)  
DE (Stellacyanin) (Phytocyanin 1).  
GN BCB OR AWI 32 OR AT5G20230 OR F5O24.120.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia, and cv. Columbia K85;  
RX MEDLINE=94124044; PubMed=8294044;  
RA van Gysel A., van Montagu M., Inze D.;  
RT "A negatively light-regulated gene from Arabidopsis thaliana encodes  
RT a protein showing high similarity to blue copper-binding proteins.";  
RL Gene 136:79-85(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Wassilewskija;  
RA Yang K.Y., Kim C.S., Cho B.H.;  
RT "Characterization of a wound-inducible Arabidopsis gene encoding a  
RT protein homologous to blue copper binding proteins.";  
RL Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.  
RN [3]

RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Landsberg erecta;  
 RX MEDLINE=20233824; PubMed=10769227;  
 RA Honma T., Goto K.;  
 RT "The Arabidopsis floral homeotic gene PISTILLATA is regulated by  
 RT discrete cis-elements responsive to induction and maintenance  
 RT signals.";  
 RL Development 127:2021-2030(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=21016721; PubMed=11130714;  
 RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,  
 RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,  
 RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,  
 RA Nakazaki N., Naruo K., Okumura S., Shinpo S., Takeuchi C., Wada T.,  
 RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,  
 RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,  
 RA Habermann K., Murray J., Johnson D., Rohlffing T., Nelson J.,  
 RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,  
 RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,  
 RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,  
 RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strowmatt C.,  
 RA Wagner-McPherson C., Wollam A., Yoakum M., Bell M., Dedhia N.,  
 RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,  
 RA Kirchhoff K., Toth K., King L., Bahret A., Miller B., Marra M.A.,  
 RA Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,  
 RA Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,  
 RA Entian K.-D., Terryn N., Hartley N., Bent E., Johnson S.,  
 RA Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,  
 RA Ramsperger U., Wedler H., Balke K., Wedler E., Peters S.,  
 RA van Staveren M., Dirkse W., Mooijman P., Klein Lankhorst R.,  
 RA Weitzenegger T., Bothe G., Rose M., Hauf J., Berneiser S., Hempel S.,  
 RA Feldpausch M., Lamberth S., Villarreal R., Gielen J., Ardiles W.,  
 RA Bents O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,  
 RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Fransz P.F.;  
 RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 408:823-826(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=22954850; PubMed=14593172;  
 RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,  
 RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,  
 RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,  
 RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,  
 RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,  
 RA Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,  
 RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,  
 RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,  
 RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,  
 RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,  
 RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,  
 RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;  
 RT "Empirical analysis of transcriptional activity in the Arabidopsis  
 RT genome.";



RL Science 302:842-846(2003).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,  
 RA Feldmann K.A.;  
 RT "Full-length cDNA from Arabidopsis thaliana.";  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.  
 RN [7]  
 RP GPI-ANCHOR.  
 RX MEDLINE=22690167; PubMed=12805588;  
 RA Borner G.H., Lilley K.S., Stevens T.J., Dupree P.;  
 RT "Identification of glycosylphosphatidylinositol-anchored proteins in  
 RT Arabidopsis. A proteomic and genomic analysis.";  
 RL Plant Physiol. 132:568-577(2003).  
 CC -!- FUNCTION: Probably acts as an electron carrier.  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -!- DEVELOPMENTAL STAGE: Maximum levels are found in 35 day old  
 CC plantlets when the rosette is mature, consisting of 8-10 fully  
 CC expanded leaves, and as the floral stem starts to form. This level  
 CC remains constant during the further life span of the plant.  
 CC -!- INDUCTION: By dark adaptation. This gives a 20-fold increase in  
 CC expression.  
 CC -!- SIMILARITY: Contains 1 plastocyanin-like domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z15058; CAA78771.1; -.  
 DR EMBL; Y18227; CAA77089.1; -.  
 DR EMBL; AB035137; BAA86999.1; -.  
 DR EMBL; AF296825; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; AY052681; AAK96585.1; -.  
 DR EMBL; AY034986; AAK59491.1; -.  
 DR EMBL; AY142577; AAN13146.1; -.  
 DR EMBL; AY088549; AAM66081.1; -.  
 DR PIR; I39698; I39698.  
 DR PIR; T51838; T51838.  
 DR HSSP; P29602; 1JER.  
 DR InterPro; IPR000923; BlueCu\_1.  
 DR InterPro; IPR008972; Cupredoxin.  
 DR InterPro; IPR003245; Plcyanin\_like.  
 DR Pfam; PF02298; Cu\_bind\_like; 1.  
 DR ProDom; PD003122; Plcyanin\_like; 1.  
 DR PROSITE; PS00196; COPPER\_BLUE; 1.  
 KW Electron transport; Membrane; Metal-binding; Copper; Signal;  
 KW Glycoprotein; GPI-anchor; Lipoprotein.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 23 174 BLUE COPPER PROTEIN.  
 FT PROPEP 175 196 REMOVED IN MATURE FORM (PROBABLE).  
 FT DOMAIN 23 118 PLASTOCYANIN-LIKE.  
 FT METAL 66 66 COPPER (BY SIMILARITY).  
 FT METAL 107 107 COPPER (BY SIMILARITY).

FT	METAL	112	112	COPPER (BY SIMILARITY).
FT	METAL	117	117	COPPER (BY SIMILARITY).
FT	DISULFID	79	113	BY SIMILARITY.
FT	CARBOHYD	98	98	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	LIPID	174	174	GPI-anchor amidated asparagine
FT				(Probable).
FT	CONFLICT	44	44	T -> S (IN REF. 1).
FT	CONFLICT	134	134	P -> L (IN REF. 1).
FT	CONFLICT	142	142	P -> L (IN REF. 1).
SQ	SEQUENCE	196 AA; 20053 MW; 05100B50518FOA56 CRC64;		

Query Match 68.3%; Score 28; DB 1; Length 196;  
 Best Local Similarity 62.5%; Pred. No. 82;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LVFFAEDF 8  
 :| ||||:  
 Db 18 VVVF AEDY 25

# RESULT 68

UT11\_ORYSA

ID UT11\_ORYSA STANDARD; PRT; 214 AA.

AC Q8S1Z1;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Probable U3 small nucleolar RNA-associated protein 11 (U3 snoRNA-associated protein 11).

GN P0468B07.17.

OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzeae; Oryza.

OX NCBI\_TaxID=4530;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Nipponbare;

RA Sasaki T., Matsumoto T., Yamamoto K.;

RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC

RT clone:P0468B07.";

RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.

CC -!- FUNCTION: Involved in nucleolar processing of pre-18S ribosomal RNA (By similarity).

CC -!- SUBUNIT: Component of the ribosomal small subunit (SSU) processome (By similarity).

CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).

CC -!- SIMILARITY: Belongs to the UTPL1 family.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AP003260; BAB89638.1; -.  
DR Gramene; Q8S1Z1; -.  
DR InterPro; IPR007144; Utp11.  
DR Pfam; PF03998; Utp11; 1.  
KW rRNA processing; Nuclear protein.  
SQ SEQUENCE 214 AA; 25321 MW; 4352134326922B43 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 214;  
Best Local Similarity 83.3%; Pred. No. 90;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 VFFAED 7  
|:||||  
Db 132 VYFAED 137

RESULT 69

Y928\_THEMEA

ID Y928\_THEMEA STANDARD; PRT; 246 AA.  
AC P56728; P46803;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Hypothetical protein TM0928.  
GN TM0928.  
OS Thermotoga maritima.  
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
OX NCBI\_TaxID=2336;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MSB8 / DSM 3109 / ATCC 43589;  
RX MEDLINE=94222023; PubMed=8168477;  
RA Darimont B., Sterner R.;  
RT "Sequence, assembly and evolution of a primordial ferredoxin from  
RT Thermotoga maritima.";  
RL EMBO J. 13:1772-1781(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MSB8 / DSM 3109 / ATCC 43589;  
RX MEDLINE=99287316; PubMed=10360571;  
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
RT genome sequence of Thermotoga maritima.";  
RL Nature 399:323-329(1999).  
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO TWO  
CC FRAMESHIFTS, ONE IN POSITIONS 229-246; THE OTHER ONE FUSES  
CC TOGETHER TM0929 AND TM0928.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U24145; AAA65436.1; ALT\_FRAME.  
DR EMBL; X82178; CAA57668.1; ALT\_FRAME.  
DR EMBL; AE001757; AAD36009.1; -.  
DR PIR; H72314; H72314.  
DR TIGR; TM0928; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 246 AA; 29365 MW; 2A32302C773042C9 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 246;  
Best Local Similarity 83.3%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 FFAEDF 8  
|| |||  
Db 217 FFEEDF 222

RESULT 70

Y128\_PYRHO

ID Y128\_PYRHO STANDARD; PRT; 253 AA.  
AC O57868;  
DT 15-MAR-2004 (Rel. 43, Created)  
DT 15-MAR-2004 (Rel. 43, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Hypothetical UPF0286 protein PH0128.  
GN PH0128.  
OS *Pyrococcus horikoshii*.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC *Pyrococcus*.  
OX NCBI\_TaxID=53953;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OT3;  
RX MEDLINE=98344137; PubMed=9679194;  
RA Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,  
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,  
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
RA Masuchi Y., Shizuya H., Kikuchi H.;  
RT "Complete sequence and gene organization of the genome of a hyper-  
RT thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3.";  
RL DNA Res. 5:55-76(1998).  
CC -!- SIMILARITY: Belongs to the UPF0286 family.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AP000001; BAA29197.1; -.  
 DR PIR; F71233; F71233.  
 DR HAMAP; MF\_00722; -; 1.  
 DR InterPro; IPR002793; DUF91.  
 DR Pfam; PF01939; DUF91; 1.  
 DR ProDom; PD013521; DUF91; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 253 AA; 29057 MW; FD601BFED45D3DCF CRC64;

Query Match 68.3%; Score 28; DB 1; Length 253;  
 Best Local Similarity 71.4%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
 || |||:  
 Db 112 VFLAEDY 118

# RESULT 71

APAH\_ACTAC

ID APAH\_ACTAC STANDARD; PRT; 275 AA.

AC O52655;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Bis(5'-nucleosyl)-tetraphosphatase, symmetrical (EC 3.6.1.41)

DE (Diadenosine tetraphosphatase) (Ap4A hydrolase) (Diadenosine 5',5'''-

DE P1,P4-tetraphosphate pyrophosphohydrolase).

GN APAH.

OS Actinobacillus actinomycetemcomitans (Haemophilus

OS actinomycetemcomitans).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

OC Pasteurellaceae; Actinobacillus.

OX NCBI\_TaxID=714;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Y4;

RA Saarela M., Fives-Taylor P.;

RL Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.

CC -!- FUNCTION: Hydrolyzes diadenosine 5',5'''-P1,P4-tetraphosphate to  
 CC yield ADP (By similarity).

CC -!- CATALYTIC ACTIVITY: P(1),P(4)-bis(5'-adenosyl)tetraphosphate +  
 CC H(2)O = 2 ADP.

CC -!- SIMILARITY: Belongs to the Ap4A hydrolase family.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AF043998; AAC00202.1; -.

DR HAMAP; MF\_00199; -; 1.

DR InterPro; IPR004617; ApaH.

DR InterPro; IPR004843; M-ppetrase.

DR InterPro; IPR006186; T\_phtase\_apaH.  
DR Pfam; PF00149; Metallophos; 1.  
DR ProDom; PD000252; T\_phtase\_apaH; 1.  
DR TIGRFAMs; TIGR00668; apaH; 1.  
KW Hydrolase.  
SQ SEQUENCE 275 AA; 31480 MW; 0843F4FE11FF6DAD CRC64;

Query Match 68.3%; Score 28; DB 1; Length 275;  
Best Local Similarity 71.4%; Pred. No. 1.2e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 VFFAEDF 8  
:| ||||  
Db 89 IFAAEDF 95

RESULT 72

Y347\_HELPJ

ID Y347\_HELPJ STANDARD; PRT; 293 AA.  
AC Q9ZMA1;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Hypothetical pseudouridine synthase JHP0321 (EC 4.2.1.70)  
DE (Pseudouridylate synthase) (Uracil hydrolyase).  
GN JHP0321.  
OS Helicobacter pylori J99 (Campylobacter pylori J99).  
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;  
OC Helicobacteraceae; Helicobacter.  
OX NCBI\_TaxID=85963;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99120557; PubMed=9923682;  
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,  
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,  
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,  
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,  
RA Trust T.J.;  
RT "Genomic sequence comparison of two unrelated isolates of the human  
RT gastric pathogen Helicobacter pylori."  
RL Nature 397:176-180(1999).  
CC -!- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine  
CC 5'-phosphate + H(2)O.  
CC -!- SIMILARITY: Belongs to the pseudouridine synthase rluA family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE001468; AAD05911.1; -.  
DR PIR; A71946; A71946.  
DR InterPro; IPR006145; PseudoU\_synth.  
DR InterPro; IPR006224; Rlu\_synth.

DR Pfam; PF00849; PseudoU\_synth\_2; 1.  
 DR ProDom; PD001819; PSI\_RLU; 1.  
 DR PROSITE; PS01129; PSI\_RLU; FALSE\_NEG.  
 KW Hypothetical protein; Lyase; Complete proteome.  
 FT ACT\_SITE 119 119 BY SIMILARITY.  
 SQ SEQUENCE 293 AA; 33668 MW; D7EF3EC54A8E6168 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 293;  
 Best Local Similarity 75.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 ||| |:  
 Db 71 LVFEAKDF 78

# RESULT 73

YC64\_AQUAE

ID YC64\_AQUAE STANDARD; PRT; 306 AA.  
 AC O67303;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein AQ\_1264.  
 GN AQ\_1264.  
 OS Aquifex aeolicus.  
 OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
 OX NCBI\_TaxID=63363;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VF5;  
 RX MEDLINE=98196666; PubMed=9537320;  
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
 RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,  
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 RT aeolicus."  
 RL Nature 392:353-358(1998).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AE000731; AAC07265.1; -.  
 DR PIR; D70409; D70409.  
 KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 13 35 POTENTIAL.  
 FT TRANSMEM 71 93 POTENTIAL.  
 FT TRANSMEM 98 120 POTENTIAL.  
 FT TRANSMEM 130 152 POTENTIAL.  
 FT TRANSMEM 165 184 POTENTIAL.  
 FT TRANSMEM 218 240 POTENTIAL.

FT TRANSMEM 247 269 POTENTIAL.  
 FT TRANSMEM 274 296 POTENTIAL.  
 SQ SEQUENCE 306 AA; 36436 MW; D9BD80ECD8DDC6D1 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 306;  
 Best Local Similarity 50.0%; Pred. No. 1.3e+02;  
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 |::| |::|  
 Db 117 LIYFTENF 124

# RESULT 74

## ALR2\_PSEAE

ID ALR2\_PSEAE STANDARD; PRT; 357 AA.  
 AC Q9HTQ2; Q9S419;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Alanine racemase, catabolic (EC 5.1.1.1).  
 GN DADX OR PA5302.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PAO1;  
 RX MEDLINE=20433225; PubMed=10977898;  
 RA Strych U., Huang H.-C., Krause K.L., Benedik M.J.;  
 RT "Characterization of the alanine racemases from Pseudomonas aeruginosa  
 RT PAO1.";  
 RL Curr. Microbiol. 41:290-294(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PAO1;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an  
 RT opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 CC -!- FUNCTION: Isomerizes L-alanine to D-alanine which is then oxidized  
 CC to pyruvate by dadA (By similarity).  
 CC -!- CATALYTIC ACTIVITY: L-alanine = D-alanine.  
 CC -!- COFACTOR: Pyridoxal phosphate (By similarity).  
 CC -!- PATHWAY: Alanine catabolism; first step.  
 CC -!- INDUCTION: By alanine.  
 CC -!- SIMILARITY: Belongs to the alanine racemase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -



CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AF165881; AAD47081.1; -.  
DR EMBL; AE004943; AAG08687.1; -.  
DR PIR; F82982; F82982.  
DR HSSP; P10724; 1BD0.  
DR HAMAP; MF\_01201; -; 1.  
DR InterPro; IPR000821; Ala\_racemase.  
DR InterPro; IPR009006; Racem\_decarbox\_C.  
DR InterPro; IPR001608; UPF0001.  
DR Pfam; PF00842; Ala\_racemase\_C; 1.  
DR Pfam; PF01168; Ala\_racemase\_N; 1.  
DR PRINTS; PR00992; ALARACEMASE.  
DR TIGRFAMs; TIGR00492; alr; 1.  
DR PROSITE; PS00395; ALANINE\_RACEMASE; 1.  
KW Isomerase; Pyridoxal phosphate; Complete proteome.  
FT ACT\_SITE 33 33 CATALYTIC BASE SPECIFIC TO D-ALANINE (BY  
FT SIMILARITY).  
FT ACT\_SITE 253 253 CATALYTIC BASE SPECIFIC TO L-ALANINE (BY  
FT SIMILARITY).  
FT BINDING 33 33 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
FT CONFLICT 138 138 S -> R (IN REF. 1).  
FT CONFLICT 178 178 A -> S (IN REF. 1).  
FT CONFLICT 262 262 S -> R (IN REF. 1).  
SQ SEQUENCE 357 AA; 38914 MW; 199F4021D02FFF48 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 357;  
Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 FFAEDF 8  
|| |||  
Db 132 FFPEDF 137

#### RESULT 75

YJGP\_HAEIN

ID YJGP\_HAEIN STANDARD; PRT; 372 AA.  
AC P45333;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical protein HI1704.  
GN HI1704.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Rd / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,

RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 RT Rd.";  
 RL Science 269:496-512(1995).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -!- SIMILARITY: STRONG, TO E.COLI YJGP.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U32843; AAC23350.1; -.  
 DR PIR; A64176; A64176.  
 DR TIGR; H11704; -.  
 DR InterPro; IPR005495; YjgP\_YjgQ.  
 DR Pfam; PF03739; YjgP\_YjgQ; 1.  
 KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 16 36 POTENTIAL.  
 FT TRANSMEM 54 74 POTENTIAL.  
 FT TRANSMEM 101 121 POTENTIAL.  
 FT TRANSMEM 266 286 POTENTIAL.  
 FT TRANSMEM 296 316 POTENTIAL.  
 FT TRANSMEM 325 345 POTENTIAL.  
 SQ SEQUENCE 372 AA; 41336 MW; 3BEFE2417366FB38 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 372;  
 Best Local Similarity 50.0%; Pred. No. 1.6e+02;  
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 |:||::|  
 Db 24 LIFFSQQF 31

# RESULT 76

RAL1\_ONCVO

ID RAL1\_ONCVO STANDARD; PRT; 388 AA.

AC P11012;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE RAL-1 protein precursor (41 kDa larval antigen).

GN RAL1.

OS Onchocerca volvulus.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;

OC Onchocercidae; Onchocerca.

OX NCBI\_TaxID=6282;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94341871; PubMed=7520419;  
 RA Rokeach L.A., Zimmerman P.A., Unnasch T.R.;  
 RT "Epitopes of the Onchocerca volvulus RAL1 antigen, a member of the  
 RT calreticulin family of proteins, recognized by sera from patients  
 RT with onchocerciasis.";  
 RL Infect. Immun. 62:3696-3704(1994).  
 RN [2]  
 RP SEQUENCE OF 53-388 FROM N.A.  
 RX MEDLINE=88273584; PubMed=2455736;  
 RA Unnasch T.R., Gallin M.Y., Soboslay P.T., Erttmann K.D., Greene B.M.;  
 RT "Isolation and characterization of expression cDNA clones encoding  
 RT antigens of Onchocerca volvulus infective larvae.";  
 RL J. Clin. Invest. 82:262-269(1988).  
 CC -!- SIMILARITY: Belongs to the calreticulin family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M20565; AAA59056.1; -.  
 DR PIR; A32507; A32507.  
 DR InterPro; IPR009033; Calret\_calnex\_P.  
 DR InterPro; IPR001580; Calreticulin.  
 DR InterPro; IPR008985; ConA\_like\_lec\_gl.  
 DR Pfam; PF00262; calreticulin; 1.  
 DR PIRSF; PIRSF002356; Calreticulin; 1.  
 DR PRINTS; PR00626; CALRETICULIN.  
 DR ProDom; PD001866; Calreticulin; 1.  
 DR PROSITE; PS00803; CALRETICULIN\_1; 1.  
 DR PROSITE; PS00804; CALRETICULIN\_2; 1.  
 DR PROSITE; PS00805; CALRETICULIN\_REPEAT; 3.  
 KW Calcium-binding; Repeat; Antigen; Signal.  
 FT SIGNAL 1 17 POTENTIAL.  
 FT CHAIN 18 388 RAL-1 PROTEIN.  
 FT DOMAIN 189 253 4 X APPROXIMATE REPEATS.  
 FT REPEAT 189 200 1-1.  
 FT REPEAT 208 219 1-2.  
 FT REPEAT 225 236 1-3.  
 FT REPEAT 242 253 1-4.  
 FT DOMAIN 257 295 3 X APPROXIMATE REPEATS.  
 FT REPEAT 257 267 2-1.  
 FT REPEAT 271 281 2-2.  
 FT REPEAT 285 295 2-3.  
 FT DOMAIN 353 388 ARG/LYS-RICH (BASIC).  
 FT DISULFID 135 161 BY SIMILARITY.  
 SQ SEQUENCE 388 AA; 45298 MW; 9537F298A2D31CD6 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 388;  
 Best Local Similarity 57.1%; Pred. No. 1.6e+02;  
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
::| |||  
Db 19 IYFKEDE 25

RESULT 77

PAN3\_HUMAN

ID PAN3\_HUMAN STANDARD; PRT; 392 AA.  
AC Q96QZ0;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Pannexin 3.  
GN PANX3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Baranova A., Ivanov D., Skoblov M., Pestova A., Kelmanson I.,  
RA Shagin D., Usman N., Lukyanov S., Panchin Y.;  
RT "Mammalian pannexin family homologous to invertebrate gap-junction  
RT proteins are differentially expressed in nervous tissue.";  
RL Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.  
CC -!- FUNCTION: STRUCTURAL COMPONENT OF THE GAP JUNCTIONS.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -!- SIMILARITY: Belongs to the innexin family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF406650; AAK95655.1; -.  
DR Genew; HGNC:20573; PANX3.  
KW Gap junction; Transmembrane.  
FT TRANSMEM 37 57 POTENTIAL.  
FT TRANSMEM 110 132 POTENTIAL.  
FT TRANSMEM 268 290 POTENTIAL.  
SQ SEQUENCE 392 AA; 44683 MW; 8988F5DCC6E1B050 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 392;  
Best Local Similarity 71.4%; Pred. No. 1.6e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
||| |:  
Db 234 VFFQEEF 240

RESULT 78

ASSY\_AQUAE

ID ASSY\_AQUAE STANDARD; PRT; 401 AA.  
 AC O67213;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Argininosuccinate synthase (EC 6.3.4.5) (Citrulline--aspartate  
 DE ligase).  
 GN ARGG OR AQ\_1140.  
 OS Aquifex aeolicus.  
 OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
 OX NCBI\_TaxID=63363;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VF5;  
 RX MEDLINE=98196666; PubMed=9537320;  
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
 RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,  
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 RT aeolicus.";  
 RL Nature 392:353-358(1998).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-citrulline + L-aspartate = AMP +  
 CC diphosphate + L-argininosuccinate.  
 CC -!- PATHWAY: Arginine biosynthesis; seventh step.  
 CC -!- SUBUNIT: Homotetramer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC -!- SIMILARITY: Belongs to the argininosuccinate synthase family.  
 CC Subfamily 1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE000725; AAC07170.1; -.  
 DR PIR; B70398; B70398.  
 DR HAMAP; MF\_00005; -; 1.  
 DR InterPro; IPR001518; Arginosuc\_synth.  
 DR Pfam; PF00764; Arginosuc\_synth; 1.  
 DR ProDom; PD003544; Arginosuc\_synth; 1.  
 DR TIGRFAMs; TIGR00032; argG; 1.  
 DR PROSITE; PS00564; ARGININOSUCCIN\_SYN\_1; 1.  
 DR PROSITE; PS00565; ARGININOSUCCIN\_SYN\_2; 1.  
 KW Arginine biosynthesis; Ligase; ATP-binding; Complete proteome.  
 SQ SEQUENCE 401 AA; 45848 MW; 2BCBA9D027CE33CF CRC64;

Query Match 68.3%; Score 28; DB 1; Length 401;  
 Best Local Similarity 62.5%; Pred. No. 1.7e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 ||::|| |  
 Db 101 LVYYAEKF 108

RESULT 79

NQRF\_HAEIN

ID NQRF\_HAEIN STANDARD; PRT; 411 AA.  
AC 005012;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Na(+)-translocating NADH-quinone reductase subunit F (EC 1.6.5.-)  
DE (Na(+)-translocating NQR subunit F) (Na(+)-NQR subunit F) (NQR complex  
DE subunit F) (NQR-1 subunit F).  
GN NQRF OR HI0171.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Rd / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
RT Rd.";  
RL Science 269:496-512(1995).  
RN [2]  
RP IDENTIFICATION AS NQR SYSTEM.  
RC STRAIN=Rd / KW20 / ATCC 51907;  
RX MEDLINE=96176316; PubMed=8601449;  
RA Hayashi M., Nakayama Y., Unemoto T.;  
RT "Existence of Na+-translocating NADH-quinone reductase in Haemophilus  
RT influenzae.";  
RL FEBS Lett. 381:174-176(1996).  
CC -!- FUNCTION: NQR complex catalyzes the reduction of ubiquinone-1 to  
CC ubiquinol by two successive reactions, coupled with the transport  
CC of Na(+) ions from the cytoplasm to the periplasm. The first step  
CC is catalyzed by nqrF, which accepts electrons from NADH and  
CC reduces ubiquinone-1 to ubisemiquinone by a one-electron transfer  
CC pathway.  
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone + Na(+) (In) = NAD(+) +  
CC ubiquinol + Na(+) (Out).  
CC -!- COFACTOR: FAD and a 2Fe-2S cluster (By similarity).  
CC -!- SUBUNIT: Composed of six subunits; nqrA, nqrB, nqrC, nqrD, nqrE  
CC and nqrF (By similarity).  
CC -!- SUBCELLULAR LOCATION: Inner membrane (Potential).  
CC -!- SIMILARITY: Belongs to the nqrF family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; U32702; AAC21841.1; -.  
DR PIR; D64052; D64052.  
DR HSSP; P23486; 1QFJ.  
DR TIGR; HI0171; -.  
DR HAMAP; MF\_00430; -, 1.  
DR InterPro; IPR006058; 2Fe2S\_fd\_BS.  
DR InterPro; IPR008333; FAD\_binding\_6.  
DR InterPro; IPR001041; Ferredoxin.  
DR InterPro; IPR001709; FPN\_cyt\_redctse.  
DR InterPro; IPR001433; Oxred\_FAD/NAD(P).  
DR Pfam; PF00970; FAD\_binding\_6; 1.  
DR Pfam; PF00111; fer2; 1.  
DR Pfam; PF00175; NAD\_binding\_1; 1.  
DR PRINTS; PR00371; FPNCR.  
DR PROSITE; PS00197; 2FE2S\_FERREDOXIN; FALSE\_NEG.  
KW Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;  
KW Flavoprotein; FAD; Metal-binding; Iron-sulfur; Iron; 2Fe-2S;  
KW Transmembrane; Inner membrane; Complete proteome.  
FT TRANSMEM 5 27 POTENTIAL.  
FT DOMAIN 38 120 FERREDOXIN.  
FT DOMAIN 276 393 CATALYTIC.  
FT METAL 73 73 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
FT METAL 79 79 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
FT METAL 82 82 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
FT METAL 114 114 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
SQ SEQUENCE 411 AA; 45705 MW; ED560053E66D23E0 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 411;  
Best Local Similarity 57.1%; Pred. No. 1.7e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
:|: |||  
Db 320 IFYQEDF 326

#### RESULT 80

##### CRT2\_BOVIN

ID CRT2\_BOVIN STANDARD; PRT; 421 AA.  
AC P42918;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Calreticulin, brain isoform 2 precursor (CRP55) (Calregulin) (HACBP).  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;

RX MEDLINE=93385184; PubMed=8373827;  
 RA Liu N., Fine R.E., Johnson R.J.;  
 RT "Comparison of cDNAs from bovine brain coding for two isoforms of  
 RT calreticulin.";  
 RL Biochim. Biophys. Acta 1202:70-76(1993).  
 CC -!- FUNCTION: This protein binds calcium. There are both high and low  
 CC affinity calcium-binding sites.  
 CC -!- SUBUNIT: Monomer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.  
 CC -!- SIMILARITY: Belongs to the calreticulin family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L13462; AAC37307.1; -.  
 DR PIR; S36799; S36799.  
 DR InterPro; IPR009033; Calret\_calnex\_P.  
 DR InterPro; IPR001580; Calreticulin.  
 DR InterPro; IPR008985; ConA\_like\_lect\_gl.  
 DR InterPro; IPR000886; ER\_target\_S.  
 DR Pfam; PF00262; calreticulin; 1.  
 DR PIRSF; PIRSF002356; Calreticulin; 1.  
 DR PRINTS; PR00626; CALRETICULIN.  
 DR ProDom; PD001866; Calreticulin; 1.  
 DR PROSITE; PS00014; ER\_TARGET; 1.  
 DR PROSITE; PS00803; CALRETICULIN\_1; 1.  
 DR PROSITE; PS00804; CALRETICULIN\_2; 1.  
 DR PROSITE; PS00805; CALRETICULIN\_REPEAT; 3.  
 KW Endoplasmic reticulum; Calcium-binding; Repeat; Signal.  
 FT SIGNAL 1 34 POTENTIAL.  
 FT CHAIN 35 421 CALRETICULIN, BRAIN ISOFORM 2.  
 FT DOMAIN 35 201 N-DOMAIN.  
 FT DOMAIN 202 312 P-DOMAIN.  
 FT DOMAIN 313 421 C-DOMAIN.  
 FT DOMAIN 195 259 4 X APPROXIMATE REPEATS.  
 FT REPEAT 195 206 1-1.  
 FT REPEAT 214 225 1-2.  
 FT REPEAT 231 242 1-3.  
 FT REPEAT 248 259 1-4.  
 FT DOMAIN 263 301 3 X APPROXIMATE REPEATS.  
 FT REPEAT 263 273 2-1.  
 FT REPEAT 277 287 2-2.  
 FT REPEAT 291 301 2-3.  
 FT DOMAIN 366 411 ASP/GLU/LYS-RICH.  
 FT DISULFID 141 167 BY SIMILARITY.  
 FT CARBOHYD 183 183 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT SITE 418 421 PREVENT SECRETION FROM ER.  
 SQ SEQUENCE 421 AA; 48812 MW; 0257E959F71528BC CRC64;

Query Match 68.3%; Score 28; DB 1; Length 421;  
 Best Local Similarity 71.4%; Pred. No. 1.8e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;



Qy 2 VFFAEDF 8  
|||:| |  
Db 38 VFFSEQF 44

RESULT 81

E1BL\_ADEM1

ID E1BL\_ADEM1 STANDARD; PRT; 433 AA.  
AC P12536;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 01-FEB-1991 (Rel. 17, Last annotation update)  
DE E1B protein, large T-antigen (55 kDa protein).  
OS Mouse adenovirus type 1 (MAV-1).  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
OX NCBI\_TaxID=10530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89012162; PubMed=3172335;  
RA Ball A.O., Williams M.E., Spindler K.R.;  
RT "Identification of mouse adenovirus type 1 early region 1: DNA  
RT sequence and a conserved transactivating function."  
RL J. Virol. 62:3947-3957(1988).

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; M22245; AAA42425.1; -.  
DR InterPro; IPR002612; AdenoE1B\_55kDa.  
DR Pfam; PF01696; Adeno\_E1B\_55K; 1.  
KW Early protein.  
SQ SEQUENCE 433 AA; 49056 MW; B67197DB24AD2D95 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 433;  
Best Local Similarity 85.7%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
|||| ||  
Db 198 VFFACDF 204

RESULT 82

NTPB\_ENTHR

ID NTPB\_ENTHR STANDARD; PRT; 458 AA.  
AC Q08637;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE V-type sodium ATP synthase subunit B (EC 3.6.3.15) (Na(+)-  
DE translocating ATPase subunit B).

GN NTPB.  
 OS Enterococcus hirae.  
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.  
 OX NCBI\_TaxID=1354;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE OF 1-25.  
 RC STRAIN=ATCC 9790;  
 RX MEDLINE=93280117; PubMed=8505293;  
 RA Takase K., Yamato I., Kakinuma Y.;  
 RT "Cloning and sequencing of the genes coding for the A and B subunits  
 RT of vacuolar-type Na(+)-ATPase from Enterococcus hirae. Coexistence of  
 RT vacuolar- and FOF1-type ATPases in one bacterial cell.";  
 RL J. Biol. Chem. 268:11610-11616(1993).  
 CC -!- FUNCTION: Involved in ATP-driven sodium extrusion.  
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.  
 CC -!- INDUCTION: By increasing intracellular Na(+) concentration.  
 CC -!- SIMILARITY: Belongs to the ATPase alpha/beta chains family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D13816; BAA02970.1; -.  
 DR EMBL; D17462; BAA04276.1; -.  
 DR PIR; B46733; B46733.  
 DR HAMAP; MF\_00310; -; 1.  
 DR InterPro; IPR000793; ATPase\_a/b\_C.  
 DR InterPro; IPR000194; ATPase\_a/bcentre.  
 DR InterPro; IPR004100; ATPase\_a/bN.  
 DR InterPro; IPR009005; F1\_ATPase\_a/bN.  
 DR Pfam; PF00006; ATP-synt\_ab; 1.  
 DR Pfam; PF00306; ATP-synt\_ab\_C; 1.  
 DR Pfam; PF02874; ATP-synt\_ab\_N; 1.  
 DR PROSITE; PS00152; ATPASE\_ALPHA\_BETA; 1.  
 KW Hydrolase; ATP synthesis; Sodium transport.  
 SQ SEQUENCE 458 AA; 51141 MW; D89D8F2F58B65262 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 458;  
 Best Local Similarity 83.3%; Pred. No. 1.9e+02;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 FFAEDF 8  
 || |||  
 Db 193 FFMEDF 198

RESULT 83  
 CRTD\_RHOSH  
 ID CRTD\_RHOSH STANDARD; PRT; 495 AA.  
 AC Q01671;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Methoxyneurosporene dehydrogenase (EC 1.14.99.-).

GN CRTD.

OS Rhodobacter sphaeroides (Rhodopseudomonas sphaeroides).

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;

OC Rhodobacteraceae; Rhodobacter.

OX NCBI\_TaxID=1063;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;

RX MEDLINE=92307398; PubMed=1612412;

RA Gari E., Toledo J.C., Gibert I., Barbe J.;

RT "Nucleotide sequence of the methoxyneurosporene dehydrogenase gene

RT from Rhodobacter sphaeroides: comparison with other bacterial

RT carotenoid dehydrogenases.";

RL FEMS Microbiol. Lett. 72:103-108(1992).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;

RX MEDLINE=95238278; PubMed=7721699;

RA Lang H.P., Cogdell R.J., Takaichi S., Hunter C.N.;

RT "Complete DNA sequence, specific Tn5 insertion map, and gene

RT assignment of the carotenoid biosynthesis pathway of Rhodobacter

RT sphaeroides.";

RL J. Bacteriol. 177:2064-2073(1995).

CC -!- FUNCTION: CONVERTS HYDROXYNEUROSPORENE TO DEMETHYLSPHEROIDENE OR

CC METHOXYNEUROSPORENE TO SPHEROIDENE.

CC -!- COFACTOR: FAD (Probable).

CC -!- PATHWAY: Carotenoid and chlorophyll biosynthesis.

CC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X63204; CAA44886.1; -.

DR EMBL; AJ010302; CAB38743.1; ALT\_SEQ.

DR PIR; S23633; S23633.

DR InterPro; IPR002937; Amino\_oxidase.

DR InterPro; IPR008150; Bac\_phytoene\_dh.

DR InterPro; IPR000205; NAD\_BS.

DR Pfam; PF01593; Amino\_oxidase; 1.

DR PROSITE; PS00982; PHYTOENE\_DH; 1.

KW Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;

KW Oxidoreductase; FAD; Flavoprotein; NAD.

FT NP\_BIND 9 42 FAD (ADP PART) (POTENTIAL).

FT CONFLICT 115 115 A -> G (IN REF. 2).

FT CONFLICT 157 157 T -> P (IN REF. 2).

FT CONFLICT 170 170 L -> M (IN REF. 2).

FT CONFLICT 273 273 L -> I (IN REF. 2).

FT CONFLICT 440 443 PHGA -> ATGP (IN REF. 1).

SQ SEQUENCE 495 AA; 52900 MW; 62227931415B253E CRC64;

Query Match

68.3%; Score 28; DB 1; Length 495;

Best Local Similarity 83.3%; Pred. No. 2.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VFFAED 7  
||||:|  
Db 331 VFFADD 336

RESULT 84

SYK\_METMP

ID SYK\_METMP STANDARD; PRT; 533 AA.  
AC O30522;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Lysyl-tRNA synthetase (EC 6.1.1.6) (Lysine--tRNA ligase) (LysRS).  
GN LYSS.  
OS Methanococcus maripaludis.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanococcaceae; Methanococcus.  
OX NCBI\_TaxID=39152;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-22.  
RC STRAIN=JJ;  
RX MEDLINE=98016282; PubMed=9353192;  
RA Ibba M., Morgan S., Curnow A.W., Pridmore D.R., Vothknecht U.C.,  
RA Gardner W., Lin W., Woese C.R., Soell D.;  
RT "A euryarchaeal lysyl-tRNA synthetase: resemblance to class I  
RT synthetases.";  
RL Science 278:1119-1122(1997).  
RN [2]  
RP CHARACTERIZATION.  
RX MEDLINE=20570460; PubMed=11121028;  
RA Soell D., Becker H.D., Plateau P., Blanquet S., Ibba M.;  
RT "Context-dependent anticodon recognition by class I lysyl-tRNA  
RT synthetases.";  
RL Proc. Natl. Acad. Sci. U.S.A. 97:14224-14228(2000).  
CC -!- CATALYTIC ACTIVITY: ATP + L-lysine + tRNA(Lys) = AMP + diphosphate  
CC + L-lysyl-tRNA(Lys).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- MISCELLANEOUS: Able to charge E.coli tRNA(Lys) in vitro.  
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF009824; AAB87410.1; -.  
DR PIR; T46975; T46975.  
DR HAMAP; MF\_00177; -; 1.  
DR InterPro; IPR002904; Lys\_tRNA-synt\_1c.  
DR InterPro; IPR008925; tRNA-synt\_bind.  
DR InterPro; IPR001412; tRNA-synt\_I.

DR Pfam; PF01921; tRNA-synt\_1f; 1.  
 DR TIGRFAMs; TIGR00467; lysS\_arch; 1.  
 DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; 1.  
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.  
 FT SITE 28 36 "HIGH" REGION.  
 FT SITE 278 282 "KMSKS" REGION.  
 FT CONFLICT 11 11 I -> L (IN REF. 1; AA SEQUENCE).  
 SQ SEQUENCE 533 AA; 61273 MW; 007FA868A57A0AC2 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 533;  
 Best Local Similarity 50.0%; Pred. No. 2.2e+02;  
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 LVFFAEDF 8  
 | : | | : | :  
 Db 59 LIFIADDY 66

# RESULT 85

CH62\_CHLCV

ID CH62\_CHLCV STANDARD; PRT; 536 AA.  
 AC P59698;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE 60 kDa chaperonin 2 (Protein Cpn60 2) (groEL protein 2).  
 GN GROL2 OR GROEL2 OR GROEL-2 OR CCA00980.  
 OS Chlamydophila caviae.  
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.  
 OX NCBI\_TaxID=83557;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GPIC;  
 RX MEDLINE=22569155; PubMed=12682364;  
 RA Read T.D., Myers G.S.A., Brunham R.C., Nelson W.C., Paulsen I.T.,  
 RA Heidelberg J., Holtzapfel E., Khouri H., Federova N.B., Carty H.A.,  
 RA Umayam L.A., Haft D.H., Peterson J., Beanan M.J., White O.,  
 RA Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G., Bavoil P.M.,  
 RA Fraser C.M.;  
 RT "Genome sequence of Chlamydophila caviae (Chlamydia psittaci GPIC):  
 RT examining the role of niche-specific genes in the evolution of the  
 RT Chlamydiaceae.";  
 RL Nucleic Acids Res. 31:2134-2147(2003).  
 CC -!- FUNCTION: Prevents misfolding and promotes the refolding and  
 CC proper assembly of unfolded polypeptides generated under stress  
 CC conditions (By similarity).  
 CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of  
 CC 7 subunits (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch).

CC

DR EMBL; AE016997; AAP05719.1; -.

DR TIGR; CCA00980; -.

DR HAMAP; MF\_00600; -; 1.

DR InterPro; IPR001844; Chaprnin\_Cpn60.

DR InterPro; IPR002423; Cpn60/TCP-1.

DR InterPro; IPR008950; GroEL-ATPase.

DR PRINTS; PR00298; CHAPERONIN60.

DR PRINTS; PR00304; TCOMPLEXTCP1.

DR PROSITE; PS00296; CHAPERONINS\_CPN60; FALSE\_NEG.

KW Chaperone; ATP-binding; Complete proteome.

SQ SEQUENCE 536 AA; 57848 MW; 053B62380116A1DD CRC64;

Query Match 68.3%; Score 28; DB 1; Length 536;

Best Local Similarity 62.5%; Pred. No. 2.2e+02;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8

|: ||||

Db 248 LIIIAEDF 255

#### RESULT 86

COX1\_RHILE

ID COX1\_RHILE STANDARD; PRT; 538 AA.

AC Q08855;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1) (Cytochrome AA3 subunit 1).

GN CTAD OR COXA.

OS Rhizobium leguminosarum.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.

OX NCBI\_TaxID=384;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=94161496; PubMed=8117073;

RA Gabel C., Bittinger M.A., Maier R.J.;

RT "Cytochrome aa3 gene regulation in members of the family

Rhizobiaceae: comparison of copper and oxygen effects in

Bradyrhizobium japonicum and Rhizobium tropici.";

RL Appl. Environ. Microbiol. 60:141-148(1994).

CC -!- FUNCTION: Cytochrome c oxidase is the component of the respiratory chain that catalyzes the reduction of oxygen to water. Subunits 1-3 form the functional core of the enzyme complex. CO I is the catalytic subunit of the enzyme. Electrons originating in cytochrome c are transferred via the copper A center of subunit 2 and heme A of subunit 1 to the bimetallic center formed by heme A3 and copper B.

CC -!- CATALYTIC ACTIVITY: 4 ferrocytochrome c + O(2) = 4 ferricytochrome c + 2 H(2)O.

CC -!- PATHWAY: Respiratory chain; terminal step.

CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.

```

CC  -!- DEVELOPMENTAL STAGE: FREE IN SOIL (NOT AS BACTEROID).
CC  -!- SIMILARITY: Belongs to the heme-copper respiratory oxidase family.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X74341; CAA52388.1; -.
DR  PIR; S36424; S36424.
DR  HSSP; P98002; 1AR1.
DR  InterPro; IPR000883; COX1.
DR  Pfam; PF00115; COX1; 1.
DR  PRINTS; PR01165; CYCOXIDASEI.
DR  PROSITE; PS00077; COX1; 1.
KW  Oxidoreductase; Heme; Copper; Transmembrane; Respiratory chain.
FT  TRANSMEM      37      57      POTENTIAL.
FT  TRANSMEM      84     104      POTENTIAL.
FT  TRANSMEM     121     141      POTENTIAL.
FT  TRANSMEM     170     190      POTENTIAL.
FT  TRANSMEM     208     228      POTENTIAL.
FT  TRANSMEM     259     279      POTENTIAL.
FT  TRANSMEM     291     311      POTENTIAL.
FT  TRANSMEM     329     349      POTENTIAL.
FT  TRANSMEM     362     382      POTENTIAL.
FT  TRANSMEM     401     421      POTENTIAL.
FT  TRANSMEM     438     458      POTENTIAL.
FT  TRANSMEM     476     496      POTENTIAL.
FT  METAL         82      82      IRON (HEME A AXIAL LIGAND) (PROBABLE).
FT  METAL        265     265      COPPER B (PROBABLE).
FT  METAL        269     269      COPPER B (PROBABLE).
FT  METAL        314     314      COPPER B (PROBABLE).
FT  METAL        315     315      COPPER B (PROBABLE).
FT  METAL        400     400      IRON (HEME A3 AXIAL LIGAND) (PROBABLE).
FT  METAL        402     402      IRON (HEME A AXIAL LIGAND) (PROBABLE).
FT  CROSSLNK      265     269      1'-histidyl-3'-tyrosine (His-Tyr)
FT                                     (By similarity).
SQ  SEQUENCE      538 AA;  59133 MW;  740C8FE6C902D9BF CRC64;

```

```

Query Match          68.3%; Score 28; DB 1; Length 538;
Best Local Similarity 62.5%; Pred. No. 2.3e+02;
Matches      5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      1 LVFFAEDF 8
          |:|| | |
Db     447 LIFFPEHF 454

```

```

RESULT 87
YABN_ECOLI
ID  YABN_ECOLI      STANDARD;      PRT;   551 AA.
AC  P33595; P75638;
DT  01-FEB-1994 (Rel. 28, Created)
DT  10-OCT-2003 (Rel. 42, Last sequence update)

```

DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical protein yabN.  
 GN YABN OR B0069.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=92334977; PubMed=1630901;  
 RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,  
 RA Isono K., Mizobuchi K., Nakata A.;  
 RT "Systematic sequencing of the Escherichia coli genome: analysis of  
 RT the 0-2.4 min region."  
 RL Nucleic Acids Res. 20:3305-3308(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12."  
 RL Science 277:1453-1474(1997).  
 CC -!- SIMILARITY: TO E.COLI YBAE.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; D10483; BAB96638.1; -.  
 DR EMBL; AE000117; AAC73180.1; -.  
 DR PIR; E64728; E64728.  
 DR EcoGene; EG12094; yabN.  
 DR InterPro; IPR000914; SBP\_bac\_5.  
 DR Pfam; PF00496; SBP\_bac\_5; 1.  
 KW Hypothetical protein; Complete proteome.  
 FT CONFLICT 356 356 V -> A (IN REF. 1).  
 FT CONFLICT 384 384 L -> F (IN REF. 1).  
 FT CONFLICT 409 409 Y -> C (IN REF. 1).  
 SQ SEQUENCE 551 AA; 63975 MW; 0F4FAA1AAFC4E731 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 551;  
 Best Local Similarity 71.4%; Pred. No. 2.3e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
 ||:|:|:  
 Db 365 LVYFAEE 371



RESULT 88

APB3\_HUMAN

ID APB3\_HUMAN STANDARD; PRT; 575 AA.  
AC O96018; O60483;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Amyloid beta A4 precursor protein-binding family A member 3 (Neuron-specific X11L2 protein) (Neuronal Munc18-1-interacting protein 3)  
DE (Mint-3) (Adapter protein X11gamma).  
GN APBA3 OR MINT3 OR X11L2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=99160409; PubMed=10049767;  
RA Tanahashi H., Tabira T.;  
RT "X11L2, a new member of X11 protein family interacts with Alzheimer's  
RT beta-amyloid precursor protein.";  
RL Biochem. Biophys. Res. Commun. 255:663-667(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,  
RA Burkhart-Schultz K., Brower A., Gordon L., Dias J., Ramirez M.,  
RA Stilwagen S., Phan H., Velasco N., Do L., Regala W., Terry A.,  
RA Garnes J., Danganan L., Erler A., Christensen M., Georgescu A.,  
RA Avila J., Liu S., Attix C., Andreise T., Trankheim M.,  
RA Amico-Keller G., Coefield J., Duarte S., Lucas S., Bruce R.,  
RA Thomas P., Quan G., Kronmiller B., Arellano A., Sanders C., Ow D.,  
RA Nolan M., Trong S., Kobayashi A., Olsen A.S., Carrano A.V.;  
RL Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.  
RN [3]  
RP SEQUENCE OF 472-575 FROM N.A.  
RX MEDLINE=99075474; PubMed=9860131;  
RA Okamoto M., Suedhof T.C.;  
RT "Mint 3: a ubiquitous mint isoform that does not bind to munc18-1 or  
RT -2.";  
RL Eur. J. Cell Biol. 77:161-165(1998).  
CC -!- FUNCTION: May modulate processing of the beta-amyloid precursor  
CC protein (APP) and hence formation of beta-APP.  
CC -!- SUBUNIT: Binds to the cytoplasmic domain of amyloid protein (APP)  
CC in vivo.  
CC -!- TISSUE SPECIFICITY: Expressed in all the tissues examined with  
CC lower levels in brain and testis.  
CC -!- DOMAIN: Composed of an N-terminal domain, a middle  
CC phosphotyrosine-binding domain (PID/PTB) that mediates binding  
CC with the cytoplasmic domain of the beta-amyloid precursor protein,  
CC and two C-terminal PDZ domains thought to attach proteins to the  
CC plasma membrane.  
CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.  
CC -!- SIMILARITY: Contains 1 PID domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AB021638; BAA74430.1; -.  
DR EMBL; AC005954; AAC72275.1; -.  
DR EMBL; AF029110; AAC17979.1; -.  
DR PIR; JG0181; JG0181.  
DR HSSP; Q02410; 1AQC.  
DR Genew; HGNC:580; APBA3.  
DR MIM; 604262; -.  
DR GO; GO:0016020; C:membrane; NAS.  
DR GO; GO:0005515; F:protein binding; IPI.  
DR GO; GO:0007242; P:intracellular signaling cascade; NAS.  
DR InterPro; IPR001478; PDZ.  
DR InterPro; IPR006020; PTB\_PID.  
DR Pfam; PF00595; PDZ; 2.  
DR Pfam; PF00640; PID; 1.  
DR SMART; SM00228; PDZ; 2.  
DR SMART; SM00462; PTB; 1.  
DR PROSITE; PS50106; PDZ; 2.  
DR PROSITE; PS01179; PID; 1.  
KW Protein transport; Repeat; Polymorphism.  
FT DOMAIN 165 171 POLY-SER.  
FT DOMAIN 217 381 PID.  
FT DOMAIN 394 480 PDZ 1.  
FT DOMAIN 485 560 PDZ 2.  
FT VARIANT 527 527 I -> F (in dbSNP:1045236).  
FT /FTId=VAR\_011822.  
FT CONFLICT 505 505 I -> IVRPRPLAPGWGGRAALSTAPEQPPPLSRAPLFLPQ  
FT (IN REF. 3).  
SQ SEQUENCE 575 AA; 61454 MW; 3B910CC74C5F3840 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 575;  
Best Local Similarity 83.3%; Pred. No. 2.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VFFAED 7  
||:||||  
Db 332 VFYAED 337

#### RESULT 89

DFA1\_ANASP

ID DFA1\_ANASP STANDARD; PRT; 576 AA.  
AC Q8YNW5;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Putative diflavin flavoprotein A 1 (EC 1.-.-.-).  
GN DFA1 OR ALL4446.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=103690;  
RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=21595285; PubMed=11759840;  
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
 RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.;  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 RT cyanobacterium *Anabaena* sp. strain PCC 7120.";  
 RL DNA Res. 8:205-213(2001).  
 CC -!- FUNCTION: Mediates electron transfer from NADH to oxygen, reducing  
 CC it to water. This modular protein has 3 redox cofactors, in other  
 CC organisms the same activity requires 2 or 3 proteins (By  
 CC similarity).  
 CC -!- MISCELLANEOUS: By homology with *norV* in *E.coli*, could be involved  
 CC in nitric oxide detoxification (By similarity).  
 CC -!- SIMILARITY: In the N-terminal section; belongs to the zinc  
 CC metallo-hydrolase family group 3.  
 CC -!- SIMILARITY: In the C-terminal section; belongs to the flavodoxin  
 CC reductase family.  
 CC -!- SIMILARITY: Contains 1 flavodoxin-like domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AP003596; BAB76145.1; -.  
 DR PIR; AF2361; AF2361.  
 DR InterPro; IPR001279; Blactmase-like.  
 DR InterPro; IPR008254; Flav\_nitox\_synth.  
 DR InterPro; IPR002563; Flavin\_Reduct.  
 DR InterPro; IPR009002; FMN\_binding.  
 DR Pfam; PF01613; Flavin\_Reduct; 1.  
 DR Pfam; PF00753; lactamase\_B; 1.  
 DR PROSITE; PS50902; FLAVODOXIN\_LIKE; 1.  
 KW Hypothetical protein; Transport; Electron transport; Oxidoreductase;  
 KW Flavoprotein; Metal-binding; Iron; Multigene family;  
 KW Complete proteome.  
 FT DOMAIN 48 240 ZINC METALLO-HYDROLASE.  
 FT DOMAIN 269 431 FLAVODOXIN-LIKE.  
 FT DOMAIN 432 576 FLAVODOXIN-REDUCTASE-LIKE.  
 FT METAL 97 97 IRON 1 (BY SIMILARITY).  
 FT METAL 99 99 IRON 1 (BY SIMILARITY).  
 FT METAL 101 101 IRON 2 (BY SIMILARITY).  
 FT METAL 164 164 IRON 1 (BY SIMILARITY).  
 FT METAL 183 183 IRON 1 AND 2 (BY SIMILARITY).  
 FT METAL 240 240 IRON 2 (BY SIMILARITY).  
 SQ SEQUENCE 576 AA; 64088 MW; AC885FEFEB8C694D CRC64;

Query Match 68.3%; Score 28; DB 1; Length 576;  
 Best Local Similarity 57.1%; Pred. No. 2.4e+02;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy            2 VFFAEDF 8  
              :|:| |:|:  
Db            271 LFYAEDY 277

RESULT 90

NEP2\_THEVU

ID    NEP2\_THEVU            STANDARD;            PRT;    585 AA.  
AC    Q08751;  
DT    01-NOV-1995 (Rel. 32, Created)  
DT    01-NOV-1995 (Rel. 32, Last sequence update)  
DT    15-MAR-2004 (Rel. 43, Last annotation update)  
DE    Neopullulanase 2 (EC 3.2.1.135) (Alpha-amylase II) (TVA II).  
GN    TVAII.  
OS    Thermoactinomyces vulgaris.  
OC    Bacteria; Firmicutes; Bacillales; Thermoactinomycetaceae;  
OC    Thermoactinomyces.  
OX    NCBI\_TaxID=2026;  
RN    [1]  
RP    SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RC    STRAIN=R-47;  
RX    MEDLINE=93222535; PubMed=7763540;  
RA    Tonozuka T., Ohtsuka M., Mogi S.-I., Sakai H., Ohta T., Sakano Y.;  
RT    "A neopullulanase-type alpha-amylase gene from Thermoactinomyces  
RT    vulgaris R-47.";  
RL    Biosci. Biotechnol. Biochem. 57:395-401(1993).  
RN    [2]  
RP    X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).  
RC    STRAIN=R-47;  
RX    MEDLINE=99241045; PubMed=10222200;  
RA    Kamitori S., Kondo S., Okuyama K., Yokota T., Shimura Y., Tonozuka T.,  
RA    Sakano Y.;  
RT    "Crystal structure of Thermoactinomyces vulgaris R-47 alpha-amylase II  
RT    (TVAII) hydrolyzing cyclodextrins and pullulan at 2.6-A resolution.";  
RL    J. Mol. Biol. 287:907-921(1999).  
RN    [3]  
RP    X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
RX    MEDLINE=22047855; PubMed=12051850;  
RA    Kamitori S., Abe A., Ohtaki A., Kaji A., Tonozuka T., Sakano Y.;  
RT    "Crystal structures and structural comparison of Thermoactinomyces  
RT    vulgaris R-47 alpha-amylase 1 (TVAI) at 1.6 A resolution and  
RT    alpha-amylase 2 (TVAII) at 2.3 A resolution.";  
RL    J. Mol. Biol. 318:443-453(2002).  
CC    -!- FUNCTION: Hydrolyzes pullulan efficiently but only a small amount  
CC            of starch. Endohydrolysis of 1,4-alpha-glucosidic linkages in  
CC            pullulan to form panose. Cleaves also (1-6)-alpha-glucosidic  
CC            linkages to form maltotriose.  
CC    -!- CATALYTIC ACTIVITY: Hydrolysis of pullulan to panose (6-alpha-D-  
CC            glucosylmaltose).  
CC    -!- COFACTOR: Binds 1 calcium ion per subunit.  
CC    -!- SUBUNIT: Monomer.  
CC    -!- SIMILARITY: Belongs to family 13 of glycosyl hydrolases.  
CC    -----  
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC    the European Bioinformatics Institute. There are no restrictions on its  
CC    use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; D13178; BAA02473.1; -.  
 DR PIR; JC1486; JC1486.  
 DR PDB; 1BVZ; 02-MAR-99.  
 DR PDB; 1G1Y; 14-MAR-01.  
 DR PDB; 1JF5; 25-DEC-02.  
 DR PDB; 1JF6; 25-DEC-02.  
 DR PDB; 1JI2; 18-DEC-02.  
 DR PDB; 1JIB; 12-AUG-03.  
 DR PDB; 1JL8; 18-DEC-02.  
 DR InterPro; IPR006589; Alp\_amyl\_cat\_sub.  
 DR InterPro; IPR006047; Alpha\_amyl\_cat.  
 DR InterPro; IPR004185; Glyco\_hydro\_13Ig.  
 DR InterPro; IPR007110; Ig-like.  
 DR Pfam; PF00128; alpha-amylase; 1.  
 DR Pfam; PF02903; alpha-amylase\_N; 1.  
 DR SMART; SM00642; Aamy; 1.  
 KW Carbohydrate metabolism; Hydrolase; Glycosidase; Calcium-binding;  
 KW 3D-structure.  
 FT ACT\_SITE 325 325  
 FT ACT\_SITE 354 354  
 FT ACT\_SITE 421 421  
 FT METAL 143 143 CALCIUM.  
 FT METAL 145 145 CALCIUM (VIA CARBONYL OXYGEN).  
 FT METAL 148 148 CALCIUM.  
 FT METAL 149 149 CALCIUM.  
 FT METAL 169 169 CALCIUM (VIA CARBONYL OXYGEN).  
 FT METAL 171 171 CALCIUM.  
 FT HELIX 3 5  
 FT STRAND 7 7  
 FT TURN 12 14  
 FT STRAND 15 19  
 FT TURN 20 21  
 FT STRAND 22 30  
 FT TURN 31 32  
 FT STRAND 36 42  
 FT TURN 44 45  
 FT TURN 48 49  
 FT STRAND 53 62  
 FT STRAND 66 74  
 FT TURN 76 77  
 FT STRAND 80 87  
 FT STRAND 93 97  
 FT TURN 98 99  
 FT STRAND 100 102  
 FT HELIX 105 108  
 FT STRAND 111 113  
 FT TURN 119 120  
 FT HELIX 126 130  
 FT STRAND 133 136  
 FT HELIX 138 141  
 FT TURN 146 147  
 FT TURN 151 152  
 FT STRAND 153 153

FT	TURN	158	159
FT	TURN	164	165
FT	STRAND	167	167
FT	HELIX	172	185
FT	TURN	186	186
FT	STRAND	189	192
FT	STRAND	196	196
FT	STRAND	212	212
FT	TURN	214	216
FT	HELIX	219	230
FT	TURN	231	233
FT	STRAND	236	240
FT	TURN	242	243
FT	STRAND	244	244
FT	TURN	247	248
FT	HELIX	250	258
FT	HELIX	259	261
FT	TURN	263	264
FT	HELIX	265	267
FT	STRAND	268	268
FT	STRAND	270	270
FT	STRAND	283	283
FT	STRAND	285	285
FT	TURN	291	292
FT	STRAND	294	296
FT	TURN	298	299
FT	HELIX	301	317
FT	TURN	318	318
FT	STRAND	321	324
FT	TURN	325	326
FT	HELIX	327	329
FT	HELIX	332	345
FT	TURN	347	348
FT	STRAND	350	353
FT	TURN	360	366
FT	STRAND	370	373
FT	HELIX	374	384
FT	TURN	385	386
FT	HELIX	391	402
FT	TURN	403	404
FT	HELIX	407	410
FT	TURN	411	412
FT	STRAND	414	416
FT	TURN	420	421
FT	HELIX	425	428
FT	TURN	429	431
FT	HELIX	433	443
FT	TURN	444	445
FT	STRAND	449	453
FT	TURN	454	455
FT	HELIX	456	458
FT	TURN	459	459
FT	TURN	465	468
FT	HELIX	476	478
FT	HELIX	481	496
FT	TURN	498	498
FT	HELIX	499	502

```

FT   STRAND      504      511
FT   TURN        512      515
FT   STRAND      516      523
FT   TURN        524      525
FT   STRAND      526      533
FT   STRAND      539      544
FT   STRAND      553      555
FT   STRAND      561      561
FT   STRAND      564      565
FT   TURN        566      567
FT   STRAND      568      573
FT   TURN        575      576
FT   STRAND      578      582
SQ   SEQUENCE    585 AA;  67467 MW;  E311813A05A7791A CRC64;

```

```

Query Match          68.3%;  Score 28;  DB 1;  Length 585;
Best Local Similarity 83.3%;  Pred. No. 2.5e+02;
Matches      5;  Conservative      0;  Mismatches      1;  Indels      0;  Gaps      0;

```

```

QY          3 FFAEDF 8
           || |||
Db         268 FFIEDF 273

```

# RESULT 91

## CRAC\_DICDI

```

ID   CRAC_DICDI      STANDARD;          PRT;   697 AA.
AC   P35401;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Protein CRAC.
GN   DAGA.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RX   MEDLINE=94375528; PubMed=8089184;
RA   Insall R., Kupsa A., Lilly P.J., Shaulsky G., Levin L.R., Loomis W.F.,
RA   Devreotes P.N.;
RT   "CRAC, a cytosolic protein containing a pleckstrin homology domain,
RT   is required for receptor and G protein-mediated activation of
RT   adenylyl cyclase in Dictyostelium.";
RL   J. Cell Biol. 126:1537-1545(1994).
RN   [2]
RP   SEQUENCE OF 1-8.
RX   MEDLINE=94245733; PubMed=8188693;
RA   Lilly P.J., Devreotes P.N.;
RT   "Identification of CRAC, a cytosolic regulator required for guanine
RT   nucleotide stimulation of adenylyl cyclase in Dictyostelium.";
RL   J. Biol. Chem. 269:14123-14129(1994).
CC   -!- FUNCTION: Couples activated G protein to adenylyl cyclase signal
CC   transduction from surface cAMP receptor.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- DEVELOPMENTAL STAGE: Tightly developmentally regulated.

```

CC    -!- SIMILARITY: Contains 1 PH domain.  
 CC    -----  
 CC    This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC    the European Bioinformatics Institute. There are no restrictions on its  
 CC    use by non-profit institutions as long as its content is in no way  
 CC    modified and this statement is not removed. Usage by and for commercial  
 CC    entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC    or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC    -----  
 DR    EMBL; U06228; AAA61782.1; -.  
 DR    PIR; A54796; A54796.  
 DR    DictyBase; DDB0001965; dagA.  
 DR    InterPro; IPR001849; PH.  
 DR    Pfam; PF00169; PH; 1.  
 DR    SMART; SM00233; PH; 1.  
 DR    PROSITE; PS50003; PH\_DOMAIN; 1.  
 FT    INIT\_MET       0       0  
 FT    DOMAIN        21      121      PH.  
 FT    DOMAIN        349      362      POLY-THR.  
 FT    DOMAIN        467      479      POLY-GLY.  
 FT    DOMAIN        612      623      POLY-SER.  
 SQ    SEQUENCE    697 AA;   78399 MW;   4B0B41AF313DEDC7 CRC64;

Query Match                   68.3%; Score 28; DB 1; Length 697;  
 Best Local Similarity   83.3%; Pred. No. 2.9e+02;  
 Matches       5; Conservative   0; Mismatches    1; Indels       0; Gaps       0;

Qy               3 FFAEDF 8  
                   || |||  
 Db              245 FFREDF 250

# RESULT 92

## YKW2\_YEAST

ID   YKW2\_YEAST       STANDARD;       PRT;   705 AA.  
 AC   P35995;  
 DT   01-JUN-1994 (Rel. 29, Created)  
 DT   01-JUN-1994 (Rel. 29, Last sequence update)  
 DT   16-OCT-2001 (Rel. 40, Last annotation update)  
 DE   Putative 82.2 kDa transcriptional regulatory protein in FRE2 5'region.  
 GN   YKL222C.  
 OS   Saccharomyces cerevisiae (Baker's yeast).  
 OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX   NCBI\_TaxID=4932;  
 RN   [1]  
 RP   SEQUENCE FROM N.A.  
 RX   MEDLINE=94378726; PubMed=8091865;  
 RA   Alexandraki D., Tzermia M.;  
 RT   "Sequencing of a 13.2 kb segment next to the left telomere of yeast  
 RT   chromosome XI revealed five open reading frames and recent  
 RT   recombination events with the right arms of chromosomes III and V."  
 RL   Yeast 10:S81-S91(1994).  
 CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC   -!- SIMILARITY: Contains 1 Zn(2)-Cys(6) fungal-type binuclear cluster  
 CC   domain.



CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X75950; CAA53551.1; -.  
 DR EMBL; Z28222; CAA82067.1; -.  
 DR PIR; S38066; S38066.  
 DR HSSP; P12351; 1PYC.  
 DR GermOnline; 139976; -.  
 DR SGD; S0001705; YKL222C.  
 DR InterPro; IPR007219; Fungal\_trans.  
 DR InterPro; IPR001138; Fungi\_TrN.  
 DR Pfam; PF04082; Fungal\_trans; 1.  
 DR Pfam; PF00172; Zn\_clus; 1.  
 DR PRINTS; PR00054; FUNGALZNCYS.  
 DR SMART; SM00066; GAL4; 1.  
 DR PROSITE; PS00463; ZN2\_CY6\_FUNGAL\_1; 1.  
 DR PROSITE; PS50048; ZN2\_CY6\_FUNGAL\_2; 1.  
 KW Hypothetical protein; Transcription regulation; DNA-binding;  
 KW Nuclear protein; Zinc; Metal-binding.  
 FT DNA\_BIND 24 52 ZN(2)-CYS(6), FUNGAL-TYPE.  
 SQ SEQUENCE 705 AA; 82248 MW; 31BF148717D72E1C CRC64;

Query Match 68.3%; Score 28; DB 1; Length 705;  
 Best Local Similarity 83.3%; Pred. No. 3e+02;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 FFAEDF 8  
 ||| ||  
 Db 201 FFASDF 206

# RESULT 93

## VIV\_ORYSA

ID VIV\_ORYSA STANDARD; PRT; 728 AA.  
 AC P37398;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Viviparous protein homolog.  
 GN VP1.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94250843; PubMed=8193305;  
 RA Hattori T., Terada T., Hamasuna S.;  
 RT "Sequence and functional analyses of the rice gene homologous to the  
 RT maize Vp1.";

RL Plant Mol. Biol. 24:805-810(1994).  
 CC -!- FUNCTION: Could participate in abscisic acid-regulated gene  
 CC expression during seed development.  
 CC -!- SIMILARITY: Contains 1 TF-B3 domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D16640; BAA04066.1; -.  
 DR TRANSFAC; T04785; -.  
 DR Gramene; P37398; -.  
 DR InterPro; IPR003340; TF\_B3.  
 DR Pfam; PF02362; B3; 1.  
 KW Developmental protein; Transcription regulation; DNA-binding;  
 KW Activator.  
 FT DOMAIN 500 685 TF-B3.  
 SQ SEQUENCE 728 AA; 75982 MW; F66882079F5FE428 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 728;  
 Best Local Similarity 83.3%; Pred. No. 3e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 VFFAED 7  
 ||||:|  
 Db 46 VFFADD 51

#### RESULT 94

##### YNN2\_YEAST

ID YNN2\_YEAST STANDARD; PRT; 1056 AA.  
 AC P53914;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical UPF0202 protein YNL132W.  
 GN YNL132W OR N1216 OR N1858.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c;  
 RX MEDLINE=96109932; PubMed=8619318;  
 RA Mallet L., Bussereau F., Jacquet M.;  
 RT "A 43.5 kb segment of yeast chromosome XIV, which contains MFA2,  
 RT MEP2, CAP/SRV2, NAM9, FKB1/FPR1/RBP1, MOM22 and CPT1, predicts an  
 RT adenosine deaminase gene and 14 new open reading frames.";  
 RL Yeast 11:1195-1209(1995).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -!- SIMILARITY: Belongs to the UPF0202 family.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; Z46843; CAA86893.1; -.  
DR EMBL; Z71408; CAA96014.1; -.  
DR PIR; S55151; S55151.  
DR GermOnline; 143138; -.  
DR SGD; S0005076; YNL132W.  
DR InterPro; IPR007807; DUF699.  
DR Pfam; PF05127; DUF699; 1.  
KW Hypothetical protein; ATP-binding; Nuclear protein.  
FT NP\_BIND 283 290 ATP (POTENTIAL).  
SQ SEQUENCE 1056 AA; 119347 MW; 76721ED0867ED618 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 1056;  
Best Local Similarity 75.0%; Pred. No. 4.4e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LVFFAEDF 8  
|| ||:||  
Db 774 LVEFAKDF 781

#### RESULT 95

##### CYA9\_HUMAN

ID CYA9\_HUMAN STANDARD; PRT; 1353 AA.  
AC O60503; O60273; Q9BWT4; Q9UGP2;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Adenylate cyclase, type IX (EC 4.6.1.1) (ATP pyrophosphate-lyase)  
DE (Adenylyl cyclase).  
GN ADCY9 OR KIAA0520.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Heart;  
RX MEDLINE=98292499; PubMed=9628827;  
RA Hacker B.M., Tomlinson J.E., Wayman G.A., Sultana R., Chan G.,  
RA Villacres E., Distèche C., Storm D.R.;  
RT "Cloning, chromosomal mapping, and regulatory properties of the human  
RT type 9 adenylyl cyclase (ADCY9).";  
RL Genomics 50:97-104(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Paterson J.M., Smith S.M., Simpson J., Grace O.C., Bell J.E.,  
RA Antoni F.A.;  
RT "Cloning and characterisation of human adenylyl cyclase IX:  
RT differential mRNA regulation and inhibition by Ca2+/calcineurin.";

RL Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Toyota T., Yamada K., Meerabux J., Hattori E., Saito K.,  
 RA Yoshitsugu K., Shimizu H., Nankai M., Toru M., Detera-Wadleigh S.D.,  
 RA Yoshikawa T.;  
 RT "Mutation screening, case control study and transmission  
 RT disequilibrium analysis of adenylyl cyclase type 9 (ADCY9) gene in  
 RT functional psychoses.";  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP SEQUENCE OF 141-1353 FROM N.A., AND REVISION TO 1154.  
 RC TISSUE=Brain;  
 RX MEDLINE=22158633; PubMed=12168954;  
 RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;  
 RT "Construction of expression-ready cDNA clones for KIAA genes: manual  
 RT curation of 330 KIAA cDNA clones.";  
 RL DNA Res. 9:99-106(2002).  
 RN [5]  
 RP SEQUENCE OF 788-1353 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=98290545; PubMed=9628581;  
 RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,  
 RA Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. IX.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.";  
 RL DNA Res. 5:31-39(1998).  
 CC -!- FUNCTION: May play a fundamental role in situations where fine  
 CC interplay between intracellular calcium and cAMP determines the  
 CC cellular function. May be a physiologically relevant docking site  
 CC for calcineurin (By similarity).  
 CC -!- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.  
 CC -!- COFACTOR: Binds 2 magnesium ions per subunit.  
 CC -!- ENZYME REGULATION: Insensitive to calcium/calmodulin, forskolin  
 CC and somatostatin. Stimulated by beta-adrenergic receptor  
 CC activation.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- DOMAIN: Composed of two homologous domains.  
 CC -!- SIMILARITY: Belongs to the adenylyl cyclase class-4/guanylyl  
 CC cyclase family.  
 CC -!- SIMILARITY: Contains 1 FKBP-type PPIase domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF036927; AAC24201.1; -.  
 DR EMBL; AJ133123; CAB65084.1; -.  
 DR EMBL; AY028959; AAK29464.1; -.  
 DR EMBL; AY028949; AAK29464.1; JOINED.  
 DR EMBL; AY028950; AAK29464.1; JOINED.  
 DR EMBL; AY028951; AAK29464.1; JOINED.

DR EMBL; AY028952; AAK29464.1; JOINED.  
 DR EMBL; AY028953; AAK29464.1; JOINED.  
 DR EMBL; AY028954; AAK29464.1; JOINED.  
 DR EMBL; AY028955; AAK29464.1; JOINED.  
 DR EMBL; AY028956; AAK29464.1; JOINED.  
 DR EMBL; AY028957; AAK29464.1; JOINED.  
 DR EMBL; AB011092; BAA25446.2; -.  
 DR HSSP; P26769; 1AB8.  
 DR Genew; HGNC:240; ADCY9.  
 DR MIM; 603302; -.  
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.  
 DR GO; GO:0004016; F:adenylate cyclase activity; TAS.  
 DR GO; GO:0007165; P:signal transduction; TAS.  
 DR InterPro; IPR001054; G\_cyclase.  
 DR Pfam; PF00211; guanylate\_cyc; 2.  
 DR SMART; SM00044; CYCc; 2.  
 DR PROSITE; PS00452; GUANYLATE\_CYCLASES\_1; 1.  
 DR PROSITE; PS50125; GUANYLATE\_CYCLASES\_2; 2.  
 KW Lyase; cAMP biosynthesis; Transmembrane; Glycoprotein; Repeat;  
 KW Metal-binding; Magnesium; Isomerase; Rotamase.  
 FT DOMAIN 1 117 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 118 138 POTENTIAL.  
 FT DOMAIN 139 141 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 142 162 POTENTIAL.  
 FT DOMAIN 163 171 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 172 192 POTENTIAL.  
 FT DOMAIN 193 215 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 216 235 POTENTIAL.  
 FT DOMAIN 236 241 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 242 259 POTENTIAL.  
 FT DOMAIN 260 280 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 281 301 POTENTIAL.  
 FT DOMAIN 302 786 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 787 807 POTENTIAL.  
 FT DOMAIN 808 818 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 819 839 POTENTIAL.  
 FT DOMAIN 840 867 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 868 888 POTENTIAL.  
 FT DOMAIN 889 891 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 892 912 POTENTIAL.  
 FT DOMAIN 913 920 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 921 941 POTENTIAL.  
 FT DOMAIN 942 975 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 976 996 POTENTIAL.  
 FT DOMAIN 997 1353 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 504 611 PPIASE, FKBP-TYPE.  
 FT METAL 399 399 MAGNESIUM 1 AND 2 (BY SIMILARITY).  
 FT METAL 400 400 MAGNESIUM 2 (VIA CARBONYL OXYGEN) (BY  
 FT SIMILARITY).  
 FT METAL 443 443 MAGNESIUM 1 AND 2 (BY SIMILARITY).  
 FT CARBOHYD 206 206 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 955 955 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 964 964 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 493 493 G -> R (IN REF. 2).  
 FT CONFLICT 884 884 V -> A (IN REF. 2).  
 FT CONFLICT 1154 1154 N -> S (IN REF. 2 AND 4).  
 FT CONFLICT 1253 1353 QHQLSISPDIRVQVDGSGRSPTDEIANLVPSVQYVDKTSI

FT GSDSSTQAKDAHLSPKRPWKEPVKAEERGRFGKAIEKDDCD  
 FT ETGIEEANELTKLNVSKSV -> APAVHLPRHPRPGGWQHR  
 FT TVSHRRDCQPGAFPCVCGQDISGF (IN REF. 1).  
 FT CONFLICT 1308 1308 P -> R (IN REF. 2).  
 SQ SEQUENCE 1353 AA; 150699 MW; 4CBF051EA49B5B7B CRC64;

Query Match 68.3%; Score 28; DB 1; Length 1353;  
 Best Local Similarity 71.4%; Pred. No. 5.7e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LVFFAED 7  
 :||| ||  
 Db 841 MVFFLED 847

# RESULT 96

## CYA9\_MOUSE

ID CYA9\_MOUSE STANDARD; PRT; 1353 AA.  
 AC P51830; Q61279;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Adenylate cyclase, type IX (EC 4.6.1.1) (ATP pyrophosphate-lyase)  
 DE (Adenylyl cyclase) (Adenylyl cyclase type 10) (ACTP10).  
 GN ADCY9.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Brain;  
 RX MEDLINE=96278831; PubMed=8662814;  
 RA Premont R.T., Matsuoka I., Mattei M.-G., Pouille Y., Defer N.,  
 RA Hanoune J.;  
 RT "Identification and characterization of a widely expressed form of  
 RT adenylyl cyclase.";  
 RL J. Biol. Chem. 271:13900-13907(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96024597; PubMed=7575502;  
 RA Paterson J.M., Smith S.M., Harmar A.J., Antoni F.A.;  
 RT "Control of a novel adenylyl cyclase by calcineurin.";  
 RL Biochem. Biophys. Res. Commun. 214:1000-1008(1995).  
 RN [3]  
 RP SEQUENCE OF 1106-1193 FROM N.A.  
 RX MEDLINE=95097788; PubMed=7528319;  
 RA Premont R.T.;  
 RT "Identification of adenylyl cyclases by amplification using  
 RT degenerate primers.";  
 RL Meth. Enzymol. 238:116-127(1994).  
 CC !- FUNCTION: May play a fundamental role in situations where fine  
 CC interplay between intracellular calcium and cAMP determines the  
 CC cellular function. May be a physiologically relevant docking site  
 CC for calcineurin.  
 CC !- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.  
 CC !- COFACTOR: Binds 2 magnesium ions per subunit.

CC    -!- ENZYME REGULATION: Insensitive to calcium/calmodulin. Stimulated  
 CC        by magnesium, forskolin and mutationally activated G protein (GS)-  
 CC        alpha. Regulated by calcineurin.  
 CC    -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC    -!- TISSUE SPECIFICITY: Found in decreasing order in skeletal muscle,  
 CC        heart, adrenal gland, ovary and brain; and to a lesser extent, in  
 CC        kidney, liver, testis, lung, thymus and spleen.  
 CC    -!- DOMAIN: Composed of two homologous domains.  
 CC    -!- SIMILARITY: Belongs to the adenylyl cyclase class-4/guanylyl  
 CC        cyclase family.  
 CC    -!- SIMILARITY: Contains 1 FKBP-type PPIase domain.

CC    -----  
 CC    This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC    the European Bioinformatics Institute. There are no restrictions on its  
 CC    use by non-profit institutions as long as its content is in no way  
 CC    modified and this statement is not removed. Usage by and for commercial  
 CC    entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC    or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC    -----

DR    EMBL; U30602; AAC52603.1; -.  
 DR    EMBL; Z50190; CAA90570.1; -.  
 DR    PIR; JC4279; JC4279.  
 DR    HSSP; P26769; 1AB8.  
 DR    MGD; MGI:108450; Adcy9.  
 DR    InterPro; IPR001054; G\_cyclase.  
 DR    Pfam; PF00211; guanylate\_cyc; 2.  
 DR    SMART; SM00044; CYCc; 2.  
 DR    PROSITE; PS00452; GUANYLATE\_CYCLASES\_1; 1.  
 DR    PROSITE; PS50125; GUANYLATE\_CYCLASES\_2; 2.  
 KW    Lyase; cAMP biosynthesis; Transmembrane; Glycoprotein; Repeat;  
 KW    Metal-binding; Magnesium; Isomerase; Rotamase.

FT	DOMAIN	1	117	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	118	138	POTENTIAL.
FT	DOMAIN	139	141	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	142	162	POTENTIAL.
FT	DOMAIN	163	171	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	172	192	POTENTIAL.
FT	DOMAIN	193	215	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	216	235	POTENTIAL.
FT	DOMAIN	236	241	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	242	259	POTENTIAL.
FT	DOMAIN	260	280	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	281	301	POTENTIAL.
FT	DOMAIN	302	786	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	787	807	POTENTIAL.
FT	DOMAIN	808	818	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	819	839	POTENTIAL.
FT	DOMAIN	840	867	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	868	888	POTENTIAL.
FT	DOMAIN	889	891	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	892	912	POTENTIAL.
FT	DOMAIN	913	920	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	921	941	POTENTIAL.
FT	DOMAIN	942	975	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	976	996	POTENTIAL.
FT	DOMAIN	997	1353	CYTOPLASMIC (POTENTIAL).

FT	DOMAIN	504	611	PPIASE, FKBP-TYPE.
FT	METAL	399	399	MAGNESIUM 1 AND 2 (BY SIMILARITY).
FT	METAL	400	400	MAGNESIUM 2 (VIA CARBONYL OXYGEN) (BY
FT				SIMILARITY).
FT	METAL	443	443	MAGNESIUM 1 AND 2 (BY SIMILARITY).
FT	CARBOHYD	206	206	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	955	955	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	964	964	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	1192	1192	N -> D (IN REF. 3).
FT	CONFLICT	1305	1305	R -> H (IN REF. 2).
SQ	SEQUENCE	1353 AA; 150953 MW; C65736A8304F689E CRC64;		

Query Match 68.3%; Score 28; DB 1; Length 1353;  
 Best Local Similarity 71.4%; Pred. No. 5.7e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
 :||| ||  
 Db 841 MVFFLED 847

# RESULT 97

## BGS2\_SCHPO

ID BGS2\_SCHPO STANDARD; PRT; 1894 AA.  
 AC O13967; Q96WS3;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 1,3-beta-glucan synthase component bgs2 (EC 2.4.1.34) (1,3-beta-D-  
 DE glucan-UDP glucosyltransferase) (Meiotic expression upregulated  
 DE protein 2l).  
 GN BGS2 OR MEU21 OR SPAC24C9.07C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,



RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of *Schizosaccharomyces pombe*."  
 RL Nature 415:871-880(2002).  
 RN [2]  
 RP SEQUENCE OF 1717-1894 FROM N.A.  
 RC STRAIN=CD16-1;  
 RX MEDLINE=21270454; PubMed=11376151;  
 RA Watanabe T., Miyashita K., Saito T.T., Yoneki T., Kakiyama Y.,  
 RA Nabeshima K., Kishi Y.A., Shimoda C., Nojima H.;  
 RT "Comprehensive isolation of meiosis-specific genes identifies novel  
 RT proteins and unusual non-coding transcripts in *Schizosaccharomyces*  
 RT *pombe*."  
 RL Nucleic Acids Res. 29:2327-2337(2001).  
 RN [3]  
 RP FUNCTION, SUBCELLULAR LOCATION, AND INDUCTION.  
 RX MEDLINE=20521933; PubMed=11069657;  
 RA Martin V., Ribas J.C., Carnero E., Duran A., Sanchez Y.;  
 RT "bgs2+, a sporulation-specific glucan synthase homologue is required  
 RT for proper ascospore wall maturation in fission yeast."  
 RL Mol. Microbiol. 38:308-321(2000).  
 RN [4]  
 RP FUNCTION, AND SUBCELLULAR LOCATION.  
 RX MEDLINE=20382733; PubMed=10922478;  
 RA Liu J., Tang X., Wang H., Balasubramanian M.;  
 RT "Bgs2p, a 1,3-beta-glucan synthase subunit, is essential for  
 RT maturation of ascospore wall in *Schizosaccharomyces pombe*."  
 RL FEBS Lett. 478:105-108(2000).  
 CC -!- FUNCTION: Has a role in ascospore development where it is required  
 CC for the assembly of a functional spore wall.  
 CC -!- CATALYTIC ACTIVITY: UDP-glucose + {(1,3)-beta-D-glucosyl}(N) = UDP  
 CC + {(1,3)-beta-D-glucosyl}(N+1).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Associated with  
 CC the forespore inner membrane.  
 CC -!- INDUCTION: By sporulation.  
 CC -!- SIMILARITY: Belongs to the glycosyltransferase family 48.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z98601; CAB11264.2; -.  
 DR EMBL; AB054305; BAB60872.1; -.  
 DR GeneDB\_SPombe; SPAC24C9.07c; -.  
 DR InterPro; IPR003440; Glyco\_trans\_48.  
 DR Pfam; PF02364; Glucan\_synthase; 1.  
 KW Transmembrane; Transferase; Glycosyltransferase; Meiosis; Sporulation.

FT TRANSMEM 566 586 POTENTIAL.  
 FT TRANSMEM 600 620 POTENTIAL.  
 FT TRANSMEM 1637 1657 POTENTIAL.  
 FT TRANSMEM 1673 1693 POTENTIAL.  
 FT TRANSMEM 1779 1799 POTENTIAL.  
 FT TRANSMEM 1837 1857 POTENTIAL.  
 SQ SEQUENCE 1894 AA; 219181 MW; 2DBA28A471AAED21 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 1894;  
 Best Local Similarity 75.0%; Pred. No. 7.9e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 | ||| ||  
 Db 1772 LNFFAADF 1779

# RESULT 98

## CA36\_CHICK

ID CA36\_CHICK STANDARD; PRT; 3137 AA.  
 AC P15989;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Collagen alpha 3(VI) chain precursor.  
 GN COL6A3.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE OF 1-853 FROM N.A.  
 RC TISSUE=Aorta;  
 RX MEDLINE=91035630; PubMed=1977751;  
 RA Doliana R., Bonaldo P., Colombatti A.;  
 RT "Multiple forms of chicken alpha 3(VI) collagen chain generated by  
 RT alternative splicing in type A repeated domains."  
 RL J. Cell Biol. 111:2197-2205(1990).  
 RN [2]  
 RP SEQUENCE OF 224-2871 FROM N.A.  
 RX MEDLINE=90212613; PubMed=2322559;  
 RA Bonaldo P., Russo V., Buccioti F., Doliana R., Colombatti A.;  
 RT "Structural and functional features of the alpha 3 chain indicate a  
 RT bridging role for chicken collagen VI in connective tissues."  
 RL Biochemistry 29:1245-1254(1990).  
 RN [3]  
 RP SEQUENCE OF 2871-3137 FROM N.A.  
 RX MEDLINE=90062147; PubMed=2584214;  
 RA Bonaldo P., Colombatti A.;  
 RT "The carboxyl terminus of the chicken alpha 3 chain of collagen VI is  
 RT a unique mosaic structure with glycoprotein Ib-like, fibronectin type  
 RT III, and Kunitz modules."  
 RL J. Biol. Chem. 264:20235-20239(1989).  
 CC -!- FUNCTION: Collagen VI acts as a cell-binding protein.  
 CC -!- SUBUNIT: Trimers composed of three different chains: alpha 1(VI),  
 CC alpha 2(VI), and alpha 3(VI).

```

CC  -!- ALTERNATIVE PRODUCTS:
CC      Event=Alternative splicing; Named isoforms=1;
CC      Comment=At least 2 isoforms are produced;
CC      Name=1;
CC      IsoId=P15989-1; Sequence=Displayed;
CC  -!- PTM: Prolines at the third position of the tripeptide repeating
CC      unit (G-X-Y) are hydroxylated in some or all of the chains.
CC  -!- SIMILARITY: Contains 1 BPTI/Kunitz inhibitor domain.
CC  -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC  -!- SIMILARITY: Contains 12 VWFA domains.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; M24282; AAA03201.1; -.
DR  PIR; A37797; A37797.
DR  HSSP; P12111; 2KNT.
DR  InterPro; IPR008161; Clg_helix.
DR  InterPro; IPR008160; Collagen.
DR  InterPro; IPR003961; FN_III.
DR  InterPro; IPR002223; Kunitz_BPTI.
DR  InterPro; IPR002035; VWF_A.
DR  Pfam; PF01391; Collagen; 6.
DR  Pfam; PF00041; fn3; 1.
DR  Pfam; PF00014; Kunitz_BPTI; 1.
DR  Pfam; PF00092; vwa; 11.
DR  PRINTS; PR00759; BASICPTASE.
DR  PRINTS; PR00453; VWFADOMAIN.
DR  ProDom; PD000007; Clg_helix; 1.
DR  ProDom; PD000222; Kunitz_BPTI; 1.
DR  SMART; SM00060; FN3; 1.
DR  SMART; SM00131; KU; 1.
DR  SMART; SM00327; VWA; 12.
DR  PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR  PROSITE; PS50279; BPTI_KUNITZ_2; 1.
DR  PROSITE; PS50234; VWFA; 12.
KW  Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW  Glycoprotein; Cell adhesion; Serine protease inhibitor; Collagen;
KW  Signal; Alternative splicing.
FT  SIGNAL      1      25      POTENTIAL.
FT  CHAIN      26     3137     COLLAGEN ALPHA 3(VI) CHAIN.
FT  DOMAIN      26     2042     NONHELICAL REGION.
FT  DOMAIN     2043     2379     TRIPLE-HELICAL REGION.
FT  DOMAIN     2380     3137     NONHELICAL REGION.
FT  DOMAIN      38      212     VWFA 1.
FT  DOMAIN     241      418     VWFA 2.
FT  DOMAIN     444      623     VWFA 3.
FT  DOMAIN     644      817     VWFA 4.
FT  DOMAIN     842     1014     VWFA 5.
FT  DOMAIN    1035     1207     VWFA 6.
FT  DOMAIN    1239     1410     VWFA 7.
FT  DOMAIN    1441     1621     VWFA 8.

```

FT	DOMAIN	1641	1814	VWFA 9.
FT	DOMAIN	1840	2029	VWFA 10.
FT	DOMAIN	2407	2587	VWFA 11.
FT	DOMAIN	2625	2821	VWFA 12.
FT	DOMAIN	2945	3043	FIBRONECTIN TYPE-III.
FT	DOMAIN	3068	3137	BPTI/KUNITZ INHIBITOR.
FT	SITE	2166	2172	INTERRUPTION IN COLLAGENOUS REGION.
FT	SITE	2254	2259	INTERRUPTION IN COLLAGENOUS REGION.
FT	SITE	2308	2309	INTERRUPTION IN COLLAGENOUS REGION.
FT	SITE	2045	2047	CELL ATTACHMENT SITE.
FT	SITE	2153	2155	CELL ATTACHMENT SITE.
FT	SITE	2159	2161	CELL ATTACHMENT SITE.
FT	ACT_SITE	3082	3083	REACTIVE BOND (BY SIMILARITY).
FT	DISULFID	3072	3122	BY SIMILARITY.
FT	DISULFID	3081	3105	BY SIMILARITY.
FT	DISULFID	3097	3118	BY SIMILARITY.
FT	CARBOHYD	201	201	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2084	2084	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2436	2436	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2563	2563	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2581	2581	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2683	2683	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2867	2867	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	3137	AA; 339595	MW; ECB428578B536357 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 3137;  
 Best Local Similarity 50.0%; Pred. No. 1.3e+03;  
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 :|:| :||  
 Db 800 MVYFMDDF 807

# RESULT 99

## MIH\_CANPG

ID MIH\_CANPG STANDARD; PRT; 78 AA.  
 AC P55846;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Molt-inhibiting hormone (MIH).  
 OS Cancer pagurus (Rock crab).  
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
 OC Eubrachyura; Cancroidea; Cancridae; Cancer.  
 OX NCBI\_TaxID=6755;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Sinus gland;  
 RX MEDLINE=97021946; PubMed=8868306;  
 RA Chung J.S., Wilkinson M.C., Webster S.G.;  
 RT "Determination of the amino acid sequence of the moult-inhibiting  
 hormone from the edible crab, Cancer pagurus."  
 RL Neuropeptides 30:95-101(1996).  
 CC -!- FUNCTION: Inhibits Y-organs where molting hormone (ecdysteroid) is  
 CC secreted. A molting cycle is initiated when MIH secretion

Q7X225

ID Q7X225 PRELIMINARY; PRT; 261 AA.  
AC Q7X225;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE GTP pyrophosphokinase.  
GN YIBM.  
OS Staphylococcus aureus (strain Mu50 / ATCC 700699).  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=158878;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Wootton M., Avison M.B., Bennett P.M., Howe R.A., MacGowan A.P.,  
RA Walsh T.R.;  
RT "Genetic analysis of seventeen genes in Staphylococcus aureus with  
RT reduced susceptibility to vancomycin (VRSA) and hetero-VRSA (hVRSA).";  
RL Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AJ564070; CAD91854.1; -.  
KW Kinase.  
SQ SEQUENCE 261 AA; 30592 MW; CE58B6DE0F263166 CRC64;

Query Match 78.0%; Score 32; DB 2; Length 261;  
Best Local Similarity 85.7%; Pred. No. 1.4e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVFFAED 7  
|||:|  
Db 158 LVFFSED 164

RESULT 41

Q7WRM0

ID Q7WRM0 PRELIMINARY; PRT; 261 AA.  
AC Q7WRM0;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE GTP pyrophosphokinase.  
GN YIBM.  
OS Staphylococcus aureus.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=1280;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Mu3, Michigan, New Jersey, Glasgow3700, Glasgow3759, PC3, St  
RC Luke, Sweden307, LIM2, LIM3, PC1, Sweden309, FranceDuf, Slovenia6096,  
RC Southampton23, LIM1, Norway1018, SMH2, LiverpoolAG, LLA, LLE,  
RC SMH11888, SMH18000, SMH18034, SMH18037, SMH12248, SMH8997, SMH10501,  
RC SMH17608, SMH17487, and SMH14017;  
RA Wootton M., Avison M.B., Bennett P.M., Howe R.A., MacGowan A.P.,  
RA Walsh T.R.;  
RT "Genetic analysis of seventeen genes in Staphylococcus aureus with  
RT reduced susceptibility to vancomycin (VRSA) and hetero-VRSA (hVRSA).";  
RL Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AJ564071; CAD91855.1; -.  
DR EMBL; AJ564072; CAD91856.1; -.

DR EMBL; AJ564073; CAD91857.1; -.  
 DR EMBL; AJ564074; CAD91858.1; -.  
 DR EMBL; AJ564075; CAD91859.1; -.  
 DR EMBL; AJ564076; CAD91860.1; -.  
 DR EMBL; AJ564077; CAD91861.1; -.  
 DR EMBL; AJ564078; CAD91862.1; -.  
 DR EMBL; AJ564079; CAD91863.1; -.  
 DR EMBL; AJ564080; CAD91864.1; -.  
 DR EMBL; AJ564081; CAD91865.1; -.  
 DR EMBL; AJ564082; CAD91866.1; -.  
 DR EMBL; AJ564083; CAD91867.1; -.  
 DR EMBL; AJ564084; CAD91868.1; -.  
 DR EMBL; AJ564085; CAD91869.1; -.  
 DR EMBL; AJ564086; CAD91870.1; -.  
 DR EMBL; AJ564087; CAD91871.1; -.  
 DR EMBL; AJ564088; CAD91872.1; -.  
 DR EMBL; AJ564089; CAD91873.1; -.  
 DR EMBL; AJ564090; CAD91874.1; -.  
 DR EMBL; AJ564091; CAD91875.1; -.  
 DR EMBL; AJ564092; CAD91876.1; -.  
 DR EMBL; AJ564093; CAD91877.1; -.  
 DR EMBL; AJ564094; CAD91878.1; -.  
 DR EMBL; AJ564095; CAD91879.1; -.  
 DR EMBL; AJ564096; CAD91880.1; -.  
 DR EMBL; AJ564097; CAD91881.1; -.  
 DR EMBL; AJ564098; CAD91882.1; -.  
 DR EMBL; AJ564099; CAD91883.1; -.  
 DR EMBL; AJ564100; CAD91884.1; -.  
 DR EMBL; AJ564101; CAD91885.1; -.  
 KW Kinase.  
 SQ SEQUENCE 261 AA; 30592 MW; CE58B6DE0F263166 CRC64;

Query Match 78.0%; Score 32; DB 2; Length 261;  
 Best Local Similarity 85.7%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
 ||||:||  
 Db 158 LVFFSED 164

#### RESULT 42

Q99V89

ID Q99V89 PRELIMINARY; PRT; 261 AA.  
 AC Q99V89;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein SA0860.  
 GN SA0860.  
 OS Staphylococcus aureus (strain N315).  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxID=158879;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21311952; PubMed=11418146;  
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,

RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Lian J., Ito T., Kanamori M.,  
 RA Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Ui Y.,  
 RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,  
 RA Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,  
 RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,  
 RA Ogasawara N., Hayashi H., Hiramatsu K.;  
 RT "Whole genome sequencing of meticillin-resistant *Staphylococcus*  
 RT aureus.";  
 RL Lancet 357:1225-1240(2001).  
 DR EMBL; AP003132; BAB42101.1; -.  
 DR PIR; B89868; B89868.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 261 AA; 30592 MW; CE58B6DE0F263166 CRC64;

Query Match 78.0%; Score 32; DB 16; Length 261;  
 Best Local Similarity 85.7%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVFFAED 7  
 ||||:|  
 Db 158 LVFFSED 164

#### RESULT 43

Q8NXD0

ID Q8NXD0 PRELIMINARY; PRT; 268 AA.  
 AC Q8NXD0;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Hypothetical protein MW0883.  
 GN MW0883.  
 OS *Staphylococcus aureus* (strain MW2).  
 OC Bacteria; Firmicutes; Bacillales; *Staphylococcus*.  
 OX NCBI\_TaxID=196620;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22040717; PubMed=12044378;  
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
 RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,  
 RA Yamamoto K., Hiramatsu K.;  
 RT "Genome and virulence determinants of high virulence community-  
 RT acquired MRSA.";  
 RL Lancet 359:1819-1827(2002).  
 DR EMBL; AP004825; BAB94748.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 268 AA; 31363 MW; F40B877860BD1140 CRC64;

Query Match 78.0%; Score 32; DB 16; Length 268;  
 Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVFFAED 7  
 ||||:|  
 Db 165 LVFFSED 171

# RESULT 44

Q83NF3

ID Q83NF3 PRELIMINARY; PRT; 379 AA.  
AC Q83NF3;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Putative integral membrane transport protein.  
GN TW575.  
OS Tropheryma whipplei (strain TW08/27) (Whipple's bacillus).  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Micrococcineae; Cellulomonadaceae; Tropheryma.  
OX NCBI\_TaxID=218496;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22495039; PubMed=12606174;  
RA Bentley S.D., Maiwald M., Murphy L.D., Pallen M.J., Yeats C.A.,  
RA Dover L.G., Norbertczak H.T., Besra G.S., Quail M.A., Harris D.E.,  
RA von Herbay A., Goble A., Rutter S., Squares R., Squares S.,  
RA Barrell B.G., Parkhill J., Relman D.A.;  
RT "Sequencing and analysis of the genome of the Whipple's disease  
RT bacterium Tropheryma whipplei.";  
RL Lancet 361:637-644(2003).  
DR EMBL; BX251412; CAD67241.1; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005215; F:transporter activity; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
DR InterPro; IPR007114; MFS.  
DR InterPro; IPR005828; Sub\_transporter.  
DR Pfam; PF00083; sugar\_tr; 1.  
DR PROSITE; PS50850; MFS; 1.  
KW Complete proteome.  
SQ SEQUENCE 379 AA; 39462 MW; 0E0F9053BACCADF2 CRC64;

Query Match 78.0%; Score 32; DB 16; Length 379;  
Best Local Similarity 75.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
| | | | | :  
Db 274 LVFFAESY 281

# RESULT 45

Q83N16

ID Q83N16 PRELIMINARY; PRT; 390 AA.  
AC Q83N16;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Multidrug efflux protein.  
GN QACA OR TWT197.  
OS Tropheryma whipplei (strain Twist) (Whipple's bacillus).  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Micrococcineae; Cellulomonadaceae; Tropheryma.  
OX NCBI\_TaxID=203267;  
RN [1]



RP SEQUENCE FROM N.A.  
 RA Raoult D., Audic S., Robert C., Ogata H., Suhre K., Drancourt M.,  
 RA Claverie J.-M.;  
 RT "Tropheryma whipplei illustrates the diversity of gene loss patterns  
 RT in small genome bacterial pathogens.";  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AE016850; AAO44294.1; -.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005215; F:transporter activity; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR007114; MFS.  
 DR InterPro; IPR005828; Sub\_transporter.  
 DR Pfam; PF00083; sugar\_tr; 1.  
 DR PROSITE; PS50850; MFS; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 390 AA; 40609 MW; 099B4A82C5AAA8B2 CRC64;

Query Match 78.0%; Score 32; DB 16; Length 390;  
 Best Local Similarity 75.0%; Pred. No. 2.1e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 ||||| :  
 Db 285 LVFFAESY 292

#### RESULT 46

O87160

ID O87160 PRELIMINARY; PRT; 398 AA.  
 AC O87160;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE WbLD protein.  
 GN WBLD.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O22;  
 RX MEDLINE=99453293; PubMed=10521656;  
 RA Yamasaki S., Shimizu T., Hoshino K., Ho S.-T., Shimada T., Nair G.B.,  
 RA Takeda Y.;  
 RT "The genes responsible for O-antigen synthesis of Vibrio cholerae O139  
 RT are closely related to those of Vibrio cholerae O22.";  
 RL Gene 237:321-332(1999).  
 DR EMBL; AB012957; BAA33635.1; -.  
 DR PIR; T44331; T44331.  
 SQ SEQUENCE 398 AA; 47080 MW; 7A1998E8C402BB74 CRC64;

Query Match 78.0%; Score 32; DB 2; Length 398;  
 Best Local Similarity 85.7%; Pred. No. 2.2e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8

|||| ||  
Db 264 VFFARDF 270

RESULT 47

Q7ULM6

ID Q7ULM6 PRELIMINARY; PRT; 656 AA.  
AC Q7ULM6;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN RB9405.  
OS Rhodopirellula baltica.  
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;  
OC Planctomycetaceae; Pirellula.  
OX NCBI\_TaxID=117;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1;  
RX MEDLINE=22735913; PubMed=12835416;  
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,  
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,  
RA Schlesner H., Amann R., Reinhardt R.;  
RT "Complete genome sequence of the marine planctomycete Pirellula sp.  
RT strain 1.";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).  
DR EMBL; BX294149; CAD76243.1; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 656 AA; 75050 MW; 977C8888D2510F02 CRC64;

Query Match 78.0%; Score 32; DB 16; Length 656;  
Best Local Similarity 75.0%; Pred. No. 3.6e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
| ||| ||  
Db 530 LAFFASDF 537

RESULT 48

Q8A6R7

ID Q8A6R7 PRELIMINARY; PRT; 1676 AA.  
AC Q8A6R7;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Conserved protein, with weak BamHI domain.  
GN BT1809.  
OS Bacteroides thetaiotaomicron.  
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;  
OC Bacteroidaceae; Bacteroides.  
OX NCBI\_TaxID=818;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VPI-5482 / ATCC 29148;  
RX MEDLINE=22550858; PubMed=12663928;

RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,  
 RA Chiang H.C., Hooper L.V., Gordon J.I.;  
 RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis."  
 RL Science 299:2074-2076(2003).  
 DR EMBL; AE016933; AAO76916.1; -.  
 DR GO; GO:0005509; F:calcium ion binding; IEA.  
 DR InterPro; IPR008938; ARM.  
 DR InterPro; IPR002048; EF-hand.  
 DR PROSITE; PS00018; EF\_HAND; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 1676 AA; 193674 MW; 28065878C0F6C961 CRC64;

Query Match 78.0%; Score 32; DB 16; Length 1676;  
 Best Local Similarity 71.4%; Pred. No. 9.4e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVFFAED 7  
 :|||  
 Db 1657 IIFFAED 1663

#### RESULT 49

Q7Z443

ID Q7Z443 PRELIMINARY; PRT; 1732 AA.  
 AC Q7Z443;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Polycystic kidney disease 1-like 3.  
 GN PKD1L3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=22666619; PubMed=12782129;  
 RA Li A., Tian X., Sung S.-W., Somlo S.;  
 RT "Identification of two novel polycystic kidney disease-1-like genes in  
 RT human and mouse genomes."  
 RL Genomics 81:596-608(2003).  
 DR EMBL; AY164485; AAO32798.1; -.  
 SQ SEQUENCE 1732 AA; 195892 MW; 64847A1C8BE5A6A1 CRC64;

Query Match 78.0%; Score 32; DB 4; Length 1732;  
 Best Local Similarity 75.0%; Pred. No. 9.7e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 | |||  
 Db 668 LTFFASDF 675

#### RESULT 50

O97917

ID O97917 PRELIMINARY; PRT; 49 AA.

AC O97917;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Amyloid protein (Fragment).  
 GN APP.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20063685; PubMed=10594237;  
 RA Konfortov B.A., Licence V.E., Miller J.R.;  
 RT "Resequencing of DNA from a diverse panel of cattle reveals a high  
 RT level of polymorphism in both intron and exon."  
 RL Mamm. Genome 10:1142-1145(1999).  
 DR EMBL; AJ133033; CAB38017.1; -.  
 DR HSSP; P05067; 1BA4.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR InterPro; IPR001255; Beta-APP.  
 DR Pfam; PF03494; Beta-APP; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 49 49  
 SQ SEQUENCE 49 AA; 5183 MW; 6287463F0559BDED CRC64;

Query Match 75.6%; Score 31; DB 6; Length 49;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VFFAED 7  
 |||||  
 Db 1 VFFAED 6

# RESULT 51

Q974D3

ID Q974D3 PRELIMINARY; PRT; 141 AA.  
 AC Q974D3;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein ST0719.  
 GN ST0719.  
 OS Sulfolobus tokodaii.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 OX NCBI\_TaxID=111955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JCM 10545 / 7;  
 RX MEDLINE=21456156; PubMed=11572479;  
 RA Kawarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
 RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,  
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
 RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,

RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,  
 RA Oshima T., Kikuchi H.;  
 RT "Complete genome sequence of an aerobic thermoacidophilic  
 RT Crenarchaeon, Sulfolobus tokodaii strain7.";  
 RL DNA Res. 8:123-140(2001).  
 DR EMBL; AP000983; BAB65727.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 141 AA; 15582 MW; BD76CE0268B9920B CRC64;

Query Match 75.6%; Score 31; DB 17; Length 141;  
 Best Local Similarity 62.5%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 |:|: |||  
 Db 22 LLFYVEDF 29

# RESULT 52

P91460

ID P91460 PRELIMINARY; PRT; 150 AA.  
 AC P91460;  
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE T20D4.15 protein.  
 GN T20D4.15.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,  
 RA Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Minx, P, Graves T.;  
 RT "The sequence of C. elegans cosmid T20D4.";  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;

RA Waterston R.;  
RL Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; U80029; AAB37582.1; -.  
DR PIR; T29939; T29939.  
DR WormPep; T20D4.15; CE13806.  
DR InterPro; IPR002542; DUF19.  
DR Pfam; PF01579; DUF19; 1.  
SQ SEQUENCE 150 AA; 17829 MW; 711200917D094EA4 CRC64;

Query Match 75.6%; Score 31; DB 5; Length 150;  
Best Local Similarity 62.5%; Pred. No. 1.3e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
||:|:|:  
Db 60 LVYFADEF 67

RESULT 53

Q8E591

ID Q8E591 PRELIMINARY; PRT; 161 AA.  
AC Q8E591;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN GBS1141.  
OS Streptococcus agalactiae (serotype III).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=216495;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NEM316 / Serotype III;  
RX MEDLINE=22242508; PubMed=12354221;  
RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,  
RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,  
RA Kunst F.;  
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing  
RT invasive neonatal disease."  
RL Mol. Microbiol. 45:1499-1513(2002).  
DR EMBL; AL766849; CAD46800.1; -.  
DR SagaList; gbs1141; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 161 AA; 18382 MW; EA300276002B122B CRC64;

Query Match 75.6%; Score 31; DB 16; Length 161;  
Best Local Similarity 75.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
| |||:|  
Db 118 LGFFAQDF 125

RESULT 54

Q82JK4

ID Q82JK4 PRELIMINARY; PRT; 179 AA.  
AC Q82JK4;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN SAV2751.  
OS Streptomyces avermitilis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=33903;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=21477403; PubMed=11572948;  
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,  
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;  
RT "Genome sequence of an industrial microorganism Streptomyces  
RT avermitilis: deducing the ability of producing secondary  
RT metabolites.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=22608306; PubMed=12692562;  
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,  
RA Sakaki Y., Hattori M., Omura S.;  
RT "Complete genome sequence and comparative analysis of the industrial  
RT microorganism Streptomyces avermitilis.";  
RL Nat. Biotechnol. 21:526-531(2003).  
DR EMBL; AP005032; BAC70462.1; -.  
DR InterPro; IPR005269; Cons\_hypoth730.  
DR Pfam; PF03641; Lysine\_decarbox; 1.  
DR TIGRFAMs; TIGR00730; TIGR00730; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 179 AA; 19286 MW; 75B53B6069659145 CRC64;

Query Match 75.6%; Score 31; DB 16; Length 179;  
Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
|||||:  
Db 158 LVFFAEE 164

# RESULT 55

Q9Z588

ID Q9Z588 PRELIMINARY; PRT; 182 AA.  
AC Q9Z588;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE SC8D9.03 protein.  
GN SCO5491 OR SC8D9.03.  
OS Streptomyces coelicolor.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Murphy L., Harris D.;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb *Streptomyces coelicolor* A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,  
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete *Streptomyces*  
 RT *coelicolor* A3(2).";  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL939123; CAB37567.1; -.  
 DR PIR; T35807; T35807.  
 DR InterPro; IPR005269; Cons\_hypoth730.  
 DR Pfam; PF03641; Lysine\_decarbox; 1.  
 DR TIGRFAMs; TIGR00730; TIGR00730; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 182 AA; 19721 MW; 8494A6107A939E49 CRC64;

Query Match 75.6%; Score 31; DB 16; Length 182;  
 Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
 |||||:  
 Db 161 LVFFAEE 167



ID Q9VQ40 PRELIMINARY; PRT; 244 AA.  
 AC Q9VQ40;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE CG17652 protein.  
 GN CG17652.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; *Drosophila*.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL; AE003585; AAF51342.1; -.  
 DR FlyBase; FBgn0031361; CG17652.  
 DR InterPro; IPR006984; DUF652.

DR Pfam; PF04900; DUF652; 1.

SQ SEQUENCE 244 AA; 27504 MW; 08D54ACD7BCCB826 CRC64;

Query Match 75.6%; Score 31; DB 5; Length 244;

Best Local Similarity 75.0%; Pred. No. 2.1e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8

||||| :|

Db 13 LVFFASNF 20

#### RESULT 57

Q8NN32

ID Q8NN32 PRELIMINARY; PRT; 259 AA.

AC Q8NN32;

DT 01-OCT-2002 (TrEMBLrel. 22, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Transcriptional regulator.

GN CGL2381.

OS Corynebacterium glutamicum (Brevibacterium flavum).

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.

OX NCBI\_TaxID=1718;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;

RA Nakagawa S.;

RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";

RL Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AP005281; BAB99774.1; -.

DR GO; GO:0003700; F:transcription factor activity; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR InterPro; IPR001647; HTH\_TetR.

DR Pfam; PF00440; tetR; 1.

KW Complete proteome.

SQ SEQUENCE 259 AA; 28363 MW; D487361648C34A2C CRC64;

Query Match 75.6%; Score 31; DB 16; Length 259;

Best Local Similarity 85.7%; Pred. No. 2.3e+02;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAED 7

| |||||

Db 37 LTFFAED 43

#### RESULT 58

Q9BGL3

ID Q9BGL3 PRELIMINARY; PRT; 262 AA.

AC Q9BGL3;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Membrane type-1 metalloproteinase (Fragment).

OS Ovis aries (Sheep).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
 OC Bovidae; Caprinae; Ovis.  
 OX NCBI\_TaxID=9940;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Ovary;  
 RA Smith G.W., Ricke W.A., Cassar C.A., Smith M.F.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF267160; AAG59848.1; -.  
 DR HSSP; P08254; 1HFS.  
 DR MEROPS; M10.014; -.  
 DR GO; GO:0005578; C:extracellular matrix; IEA.  
 DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR000585; Hemopexin.  
 DR InterPro; IPR006026; Peptidase\_M.  
 DR InterPro; IPR001818; Pept\_M10A\_M12B.  
 DR InterPro; IPR006025; Pept\_M\_Zn\_BS.  
 DR Pfam; PF00045; hemopexin; 2.  
 DR Pfam; PF00413; Peptidase\_M10; 1.  
 DR PRINTS; PR00138; MATRIXIN.  
 DR SMART; SM00120; HX; 2.  
 DR SMART; SM00235; ZnMc; 1.  
 DR PROSITE; PS00024; HEMOPEXIN; 1.  
 DR PROSITE; PS00142; ZINC PROTEASE; 1.  
 KW Hydrolase; Metal-binding; Metalloprotease; Protease; Zinc.  
 FT NON\_TER 1 1  
 FT NON\_TER 262 262  
 SQ SEQUENCE 262 AA; 29759 MW; 0C53ED138FE49271 CRC64;

Query Match 75.6%; Score 31; DB 6; Length 262;  
 Best Local Similarity 62.5%; Pred. No. 2.3e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 :||| |  
 Db 17 MIFFAEGF 24

# RESULT 59

Q8Y477

ID Q8Y477 PRELIMINARY; PRT; 291 AA.  
 AC Q8Y477;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein lmo2578.  
 GN LMO2578.  
 OS Listeria monocytogenes.  
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
 OX NCBI\_TaxID=1639;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EGD-e / Serovar 1/2a;  
 RX MEDLINE=21537279; PubMed=11679669;



Best Local Similarity 71.4%; Pred. No. 2.7e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
: ||| ||  
Db 146 IFFARDF 152

RESULT 61

O20015

ID O20015 PRELIMINARY; PRT; 361 AA.  
AC O20015;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Intron maturase (Maturase K) (Fragment).  
GN MATK.  
OS Bifora americana.  
OG Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
OC campanulids; Apiales; Apiaceae; Apioideae; apioid superclade;  
OC apioid superclade incertae sedis; Bifora.  
OX NCBI\_TaxID=48030;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Plunkett G.M., Soltis D.E., Soltis P.S.;  
RT "Evolutionary patterns in Apiaceae: inferences based on matK sequence  
RT data.";  
RL Syst. Bot. 21:477-495(1996).  
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II  
CC INTRONS (BY SIMILARITY).  
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,  
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY  
CC MITOCHONDRIAL INTRONS.  
DR EMBL; U58551; AAB66257.1; -.  
DR GO; GO:0009507; C:chloroplast; IEA.  
DR GO; GO:0006397; P:mRNA processing; IEA.  
DR InterPro; IPR002866; MatK\_N.  
DR Pfam; PF01824; MatK\_N; 1.  
KW mRNA processing; Chloroplast.  
FT NON\_TER 361 361  
SQ SEQUENCE 361 AA; 42877 MW; 017CE3453166C1C4 CRC64;

Query Match 75.6%; Score 31; DB 8; Length 361;  
Best Local Similarity 71.4%; Pred. No. 3.2e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
: ||| ||  
Db 74 IFFANDF 80

RESULT 62

O66650

ID O66650 PRELIMINARY; PRT; 368 AA.  
AC O66650;

DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein AQ\_303.  
 GN AQ\_303.  
 OS Aquifex aeolicus.  
 OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
 OX NCBI\_TaxID=63363;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VF5;  
 RX MEDLINE=98196666; PubMed=9537320;  
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
 RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,  
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 RT aeolicus.";  
 RL Nature 392:353-358(1998).  
 DR EMBL; AE000683; AAC06619.1; -.  
 DR PIR; F70327; F70327.  
 DR HSSP; O27564; 1LNQ.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0008324; F:cation transporter activity; IEA.  
 DR GO; GO:0005267; F:potassium channel activity; IEA.  
 DR GO; GO:0006813; P:potassium ion transport; IEA.  
 DR InterPro; IPR001622; K+channel\_pore.  
 DR InterPro; IPR006037; TrkAC.  
 DR InterPro; IPR003148; TrkA\_N.  
 DR Pfam; PF02080; TrkA-C; 1.  
 DR Pfam; PF02254; TrkA-N; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 368 AA; 41864 MW; FC3E0600345EB236 CRC64;

Query Match 75.6%; Score 31; DB 16; Length 368;  
 Best Local Similarity 71.4%; Pred. No. 3.2e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
 :|| |||  
 Db 107 IFFQEDF 113

# RESULT 63

Q7WD43

ID Q7WD43 PRELIMINARY; PRT; 370 AA.  
 AC Q7WD43;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Putative integral membrane protein.  
 GN BB3735.  
 OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Alcaligenaceae; Bordetella.  
 OX NCBI\_TaxID=518;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=RB50 / ATCC BAA-588;  
 RX MEDLINE=22827954; PubMed=12910271;  
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,  
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,  
 RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,  
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,  
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,  
 RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,  
 RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,  
 RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,  
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,  
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;  
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,  
 RT Bordetella parapertussis and Bordetella bronchiseptica.";  
 RL Nat. Genet. 35:32-40(2003).  
 DR EMBL; BX640448; CAE35709.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 370 AA; 40050 MW; C9E4D8123E6F288B CRC64;

Query Match 75.6%; Score 31; DB 16; Length 370;  
 Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VFFAED 7  
 |||||  
 Db 159 VFFAED 164

# RESULT 64

Q7W5K7

ID Q7W5K7 PRELIMINARY; PRT; 370 AA.  
 AC Q7W5K7;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Putative integral membrane protein.  
 GN BPP3284.  
 OS Bordetella parapertussis.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Alcaligenaceae; Bordetella.  
 OX NCBI\_TaxID=519;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=12822 / ATCC BAA-587;  
 RX MEDLINE=22827954; PubMed=12910271;  
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,  
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,  
 RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,  
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,  
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,  
 RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,  
 RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,  
 RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,  
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,  
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;  
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,  
 RT Bordetella parapertussis and Bordetella bronchiseptica.";

RL Nat. Genet. 35:32-40(2003).  
DR EMBL; BX640433; CAE38569.1; -.  
KW Complete proteome.  
SQ SEQUENCE 370 AA; 40110 MW; C9E4D8073E60823B CRC64;

Query Match 75.6%; Score 31; DB 16; Length 370;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VFFAED 7  
|||||  
Db 159 VFFAED 164

RESULT 65

Q7VW49

ID Q7VW49 PRELIMINARY; PRT; 370 AA.  
AC Q7VW49;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Putative integral membrane protein.  
GN BP2420.  
OS Bordetella pertussis.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Alcaligenaceae; Bordetella.  
OX NCBI\_TaxID=520;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;  
RX MEDLINE=22827954; PubMed=12910271;  
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,  
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,  
RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,  
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,  
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,  
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,  
RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,  
RA Rabbinoiwitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,  
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,  
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;  
RT "Comparative analysis of the genome sequences of Bordetella pertussis,  
RT Bordetella parapertussis and Bordetella bronchiseptica."  
RL Nat. Genet. 35:32-40(2003).  
DR EMBL; BX640418; CAE42691.1; -.  
KW Complete proteome.  
SQ SEQUENCE 370 AA; 40007 MW; BF3FB5A28484863F CRC64;

Query Match 75.6%; Score 31; DB 16; Length 370;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VFFAED 7  
|||||  
Db 159 VFFAED 164



RESULT 66

Q8D9D2

ID Q8D9D2 PRELIMINARY; PRT; 392 AA.  
 AC Q8D9D2;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Glycosyltransferase.  
 GN VV12671.  
 OS *Vibrio vulnificus*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; *Vibrio*.  
 OX NCBI\_TaxID=672;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CMCP6;  
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  
 RA Choy H.E.;  
 RT "Complete genome sequence of *Vibrio vulnificus* CMCP6.";  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AE016806; AA011018.1; -.  
 DR GO; GO:0004867; F:serine protease inhibitor activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR InterPro; IPR000215; Serpin.  
 DR PROSITE; PS00284; SERPIN; 1.  
 KW Transferase; Complete proteome.  
 SQ SEQUENCE 392 AA; 44970 MW; 4152383F26B0A889 CRC64;

Query Match 75.6%; Score 31; DB 16; Length 392;  
 Best Local Similarity 62.5%; Pred. No. 3.5e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 :: |||||  
 Db 4 MIVFAEDF 11

RESULT 67

O06056

ID O06056 PRELIMINARY; PRT; 438 AA.  
 AC O06056;  
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 OS *Mycobacterium gordonae*.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; *Mycobacterium*.  
 OX NCBI\_TaxID=1778;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 14470;  
 RA Picardeau M., Vincent V.;  
 RT "Identification and characterization of IS-like elements in  
 RT *Mycobacterium gordonae*.";  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; U95315; AAB54013.1; -.

KW Hypothetical protein.  
FT NON\_TER 438 438  
SQ SEQUENCE 438 AA; 47987 MW; 07E3E0125C4A62BB CRC64;

Query Match 75.6%; Score 31; DB 2; Length 438;  
Best Local Similarity 71.4%; Pred. No. 3.9e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 VFFAEDF 8  
:||||:  
Db 411 MFFAEDY 417

RESULT 68

Q8U383

ID Q8U383 PRELIMINARY; PRT; 455 AA.  
AC Q8U383;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Phospho-sugar mutase.  
GN PF0588.  
OS Pyrococcus furiosus.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Pyrococcus.  
OX NCBI\_TaxID=2261;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;  
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;  
RT "The complete sequence of the Pyrococcus furiosus genome.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AE010181; AAL80712.1; -.  
DR GO; GO:0016868; F:intramolecular transferase activity, phosph. . .; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR005841; PG\_PMM\_mutase.  
DR InterPro; IPR005844; PG\_PMM\_ABAI.  
DR InterPro; IPR005845; PG\_PMM\_ABAIL.  
DR InterPro; IPR005846; PG\_PMM\_ABAILI.  
DR InterPro; IPR005843; PG\_PMM\_C.  
DR Pfam; PF00408; PGM\_PMM; 1.  
DR Pfam; PF02878; PGM\_PMM\_I; 1.  
DR Pfam; PF02879; PGM\_PMM\_II; 1.  
DR Pfam; PF02880; PGM\_PMM\_III; 1.  
DR PRINTS; PR00509; PGMPMM.  
KW Complete proteome.  
SQ SEQUENCE 455 AA; 49632 MW; 03682D037276ABE4 CRC64;

Query Match 75.6%; Score 31; DB 17; Length 455;  
Best Local Similarity 71.4%; Pred. No. 4e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 VFFAEDF 8  
:|| |||  
Db 131 IFFKEDF 137

RESULT 69

Q8R7Q5

ID Q8R7Q5 PRELIMINARY; PRT; 519 AA.  
 AC Q8R7Q5;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE AraC-type DNA-binding domain-containing proteins.  
 GN ARAC OR TTE2343.  
 OS Thermoanaerobacter tengcongensis.  
 OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;  
 OC Thermoanaerobacteriaceae; Thermoanaerobacter.  
 OX NCBI\_TaxID=119072;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MB4 / JCM 11007;  
 RX MEDLINE=21992816; PubMed=11997336;  
 RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,  
 RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,  
 RA Tan H., Chen R., Wang J., Yu J., Yang H.;  
 RT "A complete sequence of T. tengcongensis genome."  
 RL Genome Res. 12:689-700(2002).  
 DR EMBL; AE013176; AAM25484.1; -.  
 DR GO; GO:0005622; C:intracellular; IEA.  
 DR GO; GO:0003700; F:transcription factor activity; IEA.  
 DR GO; GO:0000156; F:two-component response regulator activity; IEA.  
 DR GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.  
 DR InterPro; IPR000005; HTHAraC.  
 DR InterPro; IPR001789; Response\_reg.  
 DR Pfam; PF00165; HTH\_AraC; 2.  
 DR Pfam; PF00072; response\_reg; 1.  
 DR PRINTS; PR00032; HTHARAC.  
 DR ProDom; PD000039; Response\_reg; 1.  
 DR SMART; SM00342; HTH\_ARAC; 1.  
 DR SMART; SM00448; REC; 1.  
 DR PROSITE; PS01124; HTH\_ARAC\_FAMILY\_2; 1.  
 DR PROSITE; PS01110; RESPONSE\_REGULATORY; 1.  
 KW DNA-binding; Complete proteome.  
 SQ SEQUENCE 519 AA; 60317 MW; 9E3477C8EC53F604 CRC64;

Query Match 75.6%; Score 31; DB 16; Length 519;  
 Best Local Similarity 75.0%; Pred. No. 4.6e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 | ||||:|  
 Db 286 LKFFAEEF 293

RESULT 70

Q7X127

ID Q7X127 PRELIMINARY; PRT; 524 AA.  
 AC Q7X127;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE TrwB.

GN TRWB.  
 OS *Xanthomonas campestris* (pv. citri).  
 OG Plasmid pXcB.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xanthomonas.  
 OX NCBI\_TaxID=346;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Yuan Q., Brunings A.M., El-Yacoubi B., Shanker S., Gabriel D.W.;  
 RT "A self-mobilizing plasmid from a South American citrus canker strain  
 RT carries required pathogenicity gene pthB.";  
 RL Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AY228335; AAO72101.1; -.  
 KW Plasmid.  
 SQ SEQUENCE 524 AA; 58169 MW; AF362298C39955C2 CRC64;

Query Match 75.6%; Score 31; DB 2; Length 524;  
 Best Local Similarity 85.7%; Pred. No. 4.7e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
 | |||||  
 Db 496 VMFAEDF 502

# RESULT 71

## Q8PRK2

ID Q8PRK2 PRELIMINARY; PRT; 524 AA.  
 AC Q8PRK2;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE TrwB protein.  
 GN TRWB OR XACB0030.  
 OS *Xanthomonas axonopodis* (pv. citri).  
 OG Plasmid pXAC64.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xanthomonas.  
 OX NCBI\_TaxID=92829;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=306 / ATCC 13902 / XV 101;  
 RX MEDLINE=22022145; PubMed=12024217;  
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
 RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,  
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,  
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
 RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,

RA Setubal J.C., Kitajima J.P.;  
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
 RT host specificities."  
 RL Nature 417:459-463(2002).  
 DR EMBL; AE008925; AAM39276.1; -.  
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0009291; P:unidirectional conjugation; IEA.  
 DR InterPro; IPR003688; TRAG.  
 DR Pfam; PF02534; TRAG; 1.  
 KW Plasmid; Complete proteome.  
 SQ SEQUENCE 524 AA; 58075 MW; 17D306794D35A3D5 CRC64;

Query Match 75.6%; Score 31; DB 16; Length 524;  
 Best Local Similarity 85.7%; Pred. No. 4.7e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 VFFAEDF 8  
 | |||||  
 Db 496 VMFAEDF 502

# RESULT 72

Q96AA0

ID Q96AA0 PRELIMINARY; PRT; 537 AA.  
 AC Q96AA0;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein FLJ14869.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
 RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;  
 RT "NEDO human cDNA sequencing project."  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AK027775; BAB55358.1; -.  
 DR GO; GO:0005509; F:calcium ion binding; IEA.  
 DR InterPro; IPR000152; Asx\_hydroxyl\_S.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF00008; EGF; 6.  
 DR Pfam; PF00092; vwa; 1.  
 DR PRINTS; PR00453; VWFADOMAIN.  
 DR SMART; SM00181; EGF; 5.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 4.

DR PROSITE; PS01186; EGF\_2; 5.  
DR PROSITE; PS50234; VWFA; 1.  
KW Hypothetical protein; EGF-like domain.  
SQ SEQUENCE 537 AA; 60234 MW; 5577F5C55B389448 CRC64;

Query Match 75.6%; Score 31; DB 4; Length 537;  
Best Local Similarity 71.4%; Pred. No. 4.8e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 VFFAEDF 8  
:|:||||  
Db 395 LFYAEDF 401

# RESULT 73

## Q9XSP0

ID Q9XSP0 PRELIMINARY; PRT; 582 AA.  
AC Q9XSP0;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Membrane-type matrix metalloproteinase-1.  
GN MT1MMP.  
OS Capra hircus (Goat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Bovidae; Caprinae; Capra.  
OX NCBI\_TaxID=9925;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RA Uekiata T., Tanaka S., Mariko Y., Sato H., Seiki M., Tojo H.,  
RA Tachi C.;  
RT "cDNA sequence of membrane-type matrix metalloproteinase of Shiba goat  
RT (Capra hircus var Shiba).";  
RL Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AB010921; BAA36551.1; -.  
DR HSSP; P39900; 1JIZ.  
DR MEROPS; M10.014; -.  
DR GO; GO:0005578; C:extracellular matrix; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR000585; Hemopexin.  
DR InterPro; IPR006026; Peptidase\_M.  
DR InterPro; IPR001818; Pept\_M10A\_M12B.  
DR InterPro; IPR006025; Pept\_M\_Zn\_BS.  
DR Pfam; PF00045; hemopexin; 4.  
DR Pfam; PF00413; Peptidase\_M10; 1.  
DR Pfam; PF03933; Peptidase\_M10\_N; 1.  
DR PRINTS; PR00138; MATRIXIN.  
DR SMART; SM00120; HX; 4.  
DR SMART; SM00235; ZnMc; 1.  
DR PROSITE; PS00546; CYSTEINE\_SWITCH; 1.  
DR PROSITE; PS00024; HEMOPEXIN; 1.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.

KW Hydrolase; Membrane; Metal-binding; Metalloprotease; Protease; Zinc.  
SQ SEQUENCE 582 AA; 65878 MW; B2B4E1CEFDDB7A4BE CRC64;

Query Match 75.6%; Score 31; DB 6; Length 582;  
Best Local Similarity 62.5%; Pred. No. 5.2e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
::|||||  
Db 177 MIFFAEGF 184

RESULT 74

Q9GLE4

ID Q9GLE4 PRELIMINARY; PRT; 582 AA.  
AC Q9GLE4;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Membrane-type matrix metalloprotease 1.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Zhang B., Yan L., Moses M.A., Fang J., Miao H., Tsang P.;  
RT "Molecular cloning and biological characterization of bovine membrane-  
RT type matrix metalloprotease 1 (bMT1-MMP).";  
RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF290429; AAG28170.1; -.  
DR HSSP; P39900; 1JIZ.  
DR MEROPS; M10.014; -.  
DR GO; GO:0005578; C:extracellular matrix; IEA.  
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR000585; Hemopexin.  
DR InterPro; IPR006026; Peptidase\_M.  
DR InterPro; IPR001818; Pept\_M10A\_M12B.  
DR InterPro; IPR006025; Pept\_M\_Zn\_BS.  
DR Pfam; PF00045; hemopexin; 4.  
DR Pfam; PF00413; Peptidase\_M10; 1.  
DR Pfam; PF03933; Peptidase\_M10\_N; 1.  
DR PRINTS; PR00138; MATRIXIN.  
DR SMART; SM00120; HX; 4.  
DR SMART; SM00235; ZnMc; 1.  
DR PROSITE; PS00546; CYSTEINE\_SWITCH; 1.  
DR PROSITE; PS00024; HEMOPEXIN; 1.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Hydrolase; Metal-binding; Metalloprotease; Protease; Zinc.  
SQ SEQUENCE 582 AA; 65883 MW; 65174CE65D4040E1 CRC64;

Query Match 75.6%; Score 31; DB 6; Length 582;  
Best Local Similarity 62.5%; Pred. No. 5.2e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LVFFAEDF 8  
::|||| |  
Db 177 MIFFAEGF 184

RESULT 75

Q86C23

ID Q86C23 PRELIMINARY; PRT; 603 AA.  
AC Q86C23;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Mrell.  
OS Entamoeba histolytica.  
OC Eukaryota; Entamoebidae; Entamoeba.  
OX NCBI\_TaxID=5759;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Frisardi M., Nixon J., Loftus B., Samuelson J.;  
RT "How asexual are Giardia lamblia and Entamoeba histolytica?";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF485822; AAP35101.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR004843; M-ppestrase.  
DR Pfam; PF00149; Metallophos; 1.  
SQ SEQUENCE 603 AA; 69760 MW; CD582A1ADC96A0E1 CRC64;

Query Match 75.6%; Score 31; DB 5; Length 603;  
Best Local Similarity 85.7%; Pred. No. 5.4e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 VFFAEDF 8  
||| |||  
Db 402 VFFDEDF 408

RESULT 76

Q98SG0

ID Q98SG0 PRELIMINARY; PRT; 693 AA.  
AC Q98SG0;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Beta-amyloid precursor protein A.  
GN APP.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Van den Hurk W.H.;  
RL Thesis (2001), Department of Biological Sciences,  
RL University of Nijmegen, Nijmegen, Netherlands.  
DR EMBL; AJ298150; CAC37193.1; -.



DR HSSP; P05067; 1HZ3.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR InterPro; IPR008155; A4\_APP.  
 DR InterPro; IPR008154; A4\_extra.  
 DR InterPro; IPR001255; Beta-APP.  
 DR Pfam; PF02177; A4\_EXTRA; 1.  
 DR Pfam; PF03494; Beta-APP; 1.  
 DR PRINTS; PR00203; AMYLOIDA4.  
 DR SMART; SM00006; A4\_EXTRA; 1.  
 DR PROSITE; PS00319; A4\_EXTRA; 1.  
 DR PROSITE; PS00320; A4\_INTRA; 1.  
 KW Signal.  
 FT SIGNAL 1 18 POTENTIAL.  
 SQ SEQUENCE 693 AA; 78568 MW; CAF1DF655C1AB653 CRC64;

Query Match 75.6%; Score 31; DB 13; Length 693;  
 Best Local Similarity 85.7%; Pred. No. 6.2e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
 |||||:  
 Db 611 LVFFAEE 617

# RESULT 77

Q7Z5X1

ID Q7Z5X1 PRELIMINARY; PRT; 715 AA.  
 AC Q7Z5X1;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Ovary;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Ovary;  
 RA Strausberg R.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; BC016394; AAH16394.1; -.  
 KW Hypothetical protein.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 715 AA; 80057 MW; CBC5CEBE2D4E1F65 CRC64;

Query Match 75.6%; Score 31; DB 4; Length 715;  
 Best Local Similarity 71.4%; Pred. No. 6.4e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 VFFAEDF 8  
 :|:||||  
 Db 573 LFYAEDF 579

# RESULT 78

Q91963

ID Q91963 PRELIMINARY; PRT; 747 AA.  
 AC Q91963;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE APP747.  
 GN APP747.  
 OS Xenopus.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
 OC Xenopodinae.  
 OX NCBI\_TaxID=8353;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93129227; PubMed=1282805;  
 RA Okado H., Okamoto H.;  
 RT "A Xenopus homologue of the human beta-amyloid precursor protein:  
 RT developmental regulation of its gene expression.";  
 RL Biochem. Biophys. Res. Commun. 189:1561-1568(1992).  
 DR EMBL; S52417; AAB24853.1; -.  
 DR HSSP; P05067; 1HZ3.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0004867; F:serine protease inhibitor activity; IEA.  
 DR InterPro; IPR008155; A4\_APP.  
 DR InterPro; IPR008154; A4\_extra.  
 DR InterPro; IPR001255; Beta-APP.  
 DR InterPro; IPR002223; Kunitz\_BPTI.  
 DR Pfam; PF02177; A4\_EXTRA; 1.  
 DR Pfam; PF03494; Beta-APP; 1.  
 DR Pfam; PF00014; Kunitz\_BPTI; 1.  
 DR PRINTS; PR00203; AMYLOIDA4.  
 DR PRINTS; PR00759; BASICPTASE.  
 DR ProDom; PD000222; Kunitz\_BPTI; 1.

DR SMART; SM00131; KU; 1.  
 DR PROSITE; PS00319; A4\_EXTRA; 1.  
 DR PROSITE; PS00320; A4\_INTRA; 1.  
 DR PROSITE; PS00280; BPTI\_KUNITZ\_1; 1.  
 DR PROSITE; PS50279; BPTI\_KUNITZ\_2; 1.  
 KW Protease inhibitor; Serine protease inhibitor.  
 SQ SEQUENCE 747 AA; 84893 MW; A75E81885681D948 CRC64;

Query Match 75.6%; Score 31; DB 13; Length 747;  
 Best Local Similarity 85.7%; Pred. No. 6.7e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
 |||||:  
 Db 665 LVFFAEE 671

# RESULT 79

O77066

ID O77066 PRELIMINARY; PRT; 748 AA.  
 AC O77066;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Diapause associated protein 2.  
 GN DAP2.  
 OS Choristoneura fumiferana (Spruce budworm).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
 OC Tortricodea; Tortricidae; Tortricinae; Choristoneura.  
 OX NCBI\_TaxID=7141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Palli S.R., Ladd T.R., Ricci A.R., Primavera M., Mungrue I.N.,  
 RA Pang A.S.D., Retnakaran A.;  
 RT "Synthesis of the same two proteins prior to larval diapause and  
 RT pupation in the spruce budworm, Choristoneura fumiferana.";  
 RL J. Insect Physiol. 44:509-524(1998).  
 DR EMBL; AF007768; AAC35429.1; -.  
 DR HSSP; P04253; 1LLA.  
 DR GO; GO:0005344; F:oxygen transporter activity; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR008922; Di-copper\_centre.  
 DR InterPro; IPR000896; Hemocyanin.  
 DR InterPro; IPR005203; hemocyanin\_C.  
 DR InterPro; IPR005204; hemocyanin\_N.  
 DR InterPro; IPR007110; Ig-like.  
 DR Pfam; PF00372; hemocyanin; 1.  
 DR Pfam; PF03723; hemocyanin\_C; 1.  
 DR Pfam; PF03722; hemocyanin\_N; 1.  
 DR PRINTS; PR00187; HAEMOCYANIN.  
 DR PROSITE; PS00209; HEMOCYANIN\_1; 1.  
 DR PROSITE; PS00210; HEMOCYANIN\_2; 1.  
 SQ SEQUENCE 748 AA; 88192 MW; 7354407BCF285CF5 CRC64;

Query Match 75.6%; Score 31; DB 5; Length 748;  
 Best Local Similarity 62.5%; Pred. No. 6.7e+02;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
|::||:|  
Db 109 LLYFAKDF 116

RESULT 80

Q96FT5

ID Q96FT5 PRELIMINARY; PRT; 937 AA.  
AC Q96FT5;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Matrilin 2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; BC010444; AAH10444.1; -.  
DR GO; GO:0005509; F:calcium ion binding; IEA.  
DR InterPro; IPR000152; Asx\_hydroxyl\_S.  
DR InterPro; IPR001881; EGF\_Ca.  
DR InterPro; IPR006209; EGF\_like.  
DR InterPro; IPR002035; VWF\_A.  
DR Pfam; PF00008; EGF; 10.  
DR Pfam; PF00092; vwa; 2.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00179; EGF\_CA; 1.  
DR SMART; SM00327; VWA; 2.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 9.  
DR PROSITE; PS01186; EGF\_2; 9.  
DR PROSITE; PS50234; VWFA; 2.  
KW EGF-like domain.  
SQ SEQUENCE 937 AA; 104775 MW; 7C76C5CE70F06E40 CRC64;

Query Match 75.6%; Score 31; DB 4; Length 937;  
Best Local Similarity 71.4%; Pred. No. 8.4e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
:|:||||  
Db 814 LFYAEDF 820

RESULT 81

Q99K64

ID Q99K64 PRELIMINARY; PRT; 956 AA.  
AC Q99K64;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Matrilin 2.  
 GN MATN2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; BC005429; AAH05429.1; -.  
 DR HSSP; P05099; 1AQ5.  
 DR MGD; MGI:109613; Matn2.  
 DR GO; GO:0005578; C:extracellular matrix; IDA.  
 DR InterPro; IPR000152; Asx\_hydroxyl\_S.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF00008; EGF; 10.  
 DR Pfam; PF00092; vwa; 2.  
 DR PRINTS; PR00453; VWFADOMAIN.  
 DR SMART; SM00181; EGF; 10.  
 DR SMART; SM00327; VWA; 2.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 7.  
 DR PROSITE; PS01186; EGF\_2; 9.  
 DR PROSITE; PS50234; VWFA; 2.  
 KW EGF-like domain.  
 SQ SEQUENCE 956 AA; 106706 MW; 373BFC774370BFEE CRC64;

Query Match 75.6%; Score 31; DB 11; Length 956;  
 Best Local Similarity 71.4%; Pred. No. 8.6e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
 :|:||||  
 Db 814 LFYAEDF 820

# RESULT 82

Q8R542

ID Q8R542 PRELIMINARY; PRT; 956 AA.  
 AC Q8R542;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Matrilin-2.  
 GN MATN2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/Sv;  
 RX MEDLINE=21841526; PubMed=11852232;  
 RA Mates L., Korpos E., Deak F., Liu Z., Beier D.R., Aszodi A., Kiss I.;

RT "Comparative analysis of the mouse and human genes (Matn2 and MATN2)  
RT for matrilin-2, a filament-forming protein widely distributed in  
RT extracellular matrices.";  
RL Matrix Biol. 21:163-174(2002).  
DR EMBL; AF358844; AAM11539.1; -.  
DR EMBL; AF358831; AAM11539.1; JOINED.  
DR EMBL; AF358832; AAM11539.1; JOINED.  
DR EMBL; AF358833; AAM11539.1; JOINED.  
DR EMBL; AF358834; AAM11539.1; JOINED.  
DR EMBL; AF358835; AAM11539.1; JOINED.  
DR EMBL; AF358836; AAM11539.1; JOINED.  
DR EMBL; AF358837; AAM11539.1; JOINED.  
DR EMBL; AF358838; AAM11539.1; JOINED.  
DR EMBL; AF358839; AAM11539.1; JOINED.  
DR EMBL; AF358840; AAM11539.1; JOINED.  
DR EMBL; AF358841; AAM11539.1; JOINED.  
DR EMBL; AF358842; AAM11539.1; JOINED.  
DR EMBL; AF358843; AAM11539.1; JOINED.  
DR GO; GO:0005509; F:calcium ion binding; IEA.  
DR InterPro; IPR000152; Asx\_hydroxyl\_S.  
DR InterPro; IPR001881; EGF\_Ca.  
DR InterPro; IPR006209; EGF\_like.  
DR InterPro; IPR006210; IEGF.  
DR InterPro; IPR002035; VWF\_A.  
DR Pfam; PF00008; EGF; 10.  
DR Pfam; PF00092; vwa; 2.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00181; EGF; 10.  
DR SMART; SM00327; VWA; 2.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 7.  
DR PROSITE; PS01186; EGF\_2; 9.  
DR PROSITE; PS50234; VWFA; 2.  
KW EGF-like domain.  
SQ SEQUENCE 956 AA; 106748 MW; 3E4A608FF92BDE55 CRC64;

Query Match 75.6%; Score 31; DB 11; Length 956;  
Best Local Similarity 71.4%; Pred. No. 8.6e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
:|:||||  
Db 814 LFYAEDF 820

# RESULT 83

## Q8NDE6

ID Q8NDE6 PRELIMINARY; PRT; 1016 AA.  
AC Q8NDE6;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein (Fragment).  
GN DKFZP434I245.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Duesterhoeft A., Lauber J., Mewes H.W., Weil B., Wiemann S.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL833931; CAD38787.1; -.  
 DR GO; GO:0005509; F:calcium ion binding; IEA.  
 DR InterPro; IPR000152; Asx\_hydroxyl\_S.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF00008; EGF; 10.  
 DR Pfam; PF00092; vwa; 2.  
 DR PRINTS; PR00453; VWFADOMAIN.  
 DR SMART; SM00181; EGF; 10.  
 DR SMART; SM00179; EGF\_CA; 10.  
 DR SMART; SM00327; VWA; 2.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 9.  
 DR PROSITE; PS01186; EGF\_2; 9.  
 DR PROSITE; PS50234; VWFA; 2.  
 KW Hypothetical protein.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 1016 AA; 113088 MW; 4BDC0DE391DA7C82 CRC64;

Query Match 75.6%; Score 31; DB 4; Length 1016;  
 Best Local Similarity 71.4%; Pred. No. 9.2e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
 :|:||||  
 Db 874 LFYAEDF 880

#### RESULT 84

##### Q8WRY3

ID Q8WRY3 PRELIMINARY; PRT; 1296 AA.  
 AC Q8WRY3;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Cytoadherence linked asexual protein-like protein.  
 OS Plasmodium chabaudi.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5825;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AS;  
 RA Holt D.C., Fischer K., Tchavtchitch M., Wilson D.W., Hauquitz N.E.,  
 RA Hawthorne P.L., Gardiner D.L., Trenholme K.R., Kemp D.J.;  
 RT "Clags in *P. falciparum* and other species of *Plasmodium*."  
 RL Mol. Biochem. Parasitol. 118:259-263(2002).  
 DR EMBL; AF387740; AAL57302.1; -.  
 DR InterPro; IPR005553; CLAG.  
 DR InterPro; IPR001202; WW\_Rsp5\_WWP.  
 DR Pfam; PF03805; CLAG; 1.  
 DR PROSITE; PS01159; WW\_DOMAIN\_1; 1.

SQ SEQUENCE 1296 AA; 152409 MW; 35B134DF9A54EDAD CRC64;

Query Match 75.6%; Score 31; DB 5; Length 1296;  
Best Local Similarity 62.5%; Pred. No. 1.2e+03;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LVFFAEDF 8  
::|||||  
Db 790 IIFFAEPF 797

RESULT 85

Q8NIT0

ID Q8NIT0 PRELIMINARY; PRT; 1364 AA.  
AC Q8NIT0;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Probable xanthine dehydrogenase.  
GN B23E9.040.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,  
RA Nyakatura G., Mewes H.W., Mannhaupt G.;  
RL Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA German Neurospora genome project;  
RL Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AL807372; CAD37030.1; -.  
DR PIR; T51920; T51920.  
DR GO; GO:0005489; F:electron transporter activity; IEA.  
DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR002888; 2Fe-2S\_bind.  
DR InterPro; IPR006058; 2Fe2S\_fd\_BS.  
DR InterPro; IPR008274; Aldxan\_dh\_bind.  
DR InterPro; IPR000674; Aldxan\_dh\_hamm.  
DR InterPro; IPR005107; CO\_deh\_flav\_C.  
DR InterPro; IPR002346; dehydrog\_molyb.  
DR InterPro; IPR001041; Ferredoxin.  
DR Pfam; PF01315; Ald\_Xan\_dh\_C; 1.  
DR Pfam; PF02738; Ald\_Xan\_dh\_C2; 1.  
DR Pfam; PF03450; CO\_deh\_flav\_C; 1.  
DR Pfam; PF00941; FAD\_binding\_5; 1.  
DR Pfam; PF00111; fer2; 1.  
DR Pfam; PF01799; fer2\_2; 1.  
DR ProDom; PD186071; 2Fe-2S\_bind; 1.  
DR PROSITE; PS00197; 2FE2S\_FERREDOXIN; 1.  
KW Iron; Iron-sulfur.  
SQ SEQUENCE 1364 AA; 149832 MW; 282B3187A22D2127 CRC64;

Query Match 75.6%; Score 31; DB 3; Length 1364;



Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VFFAED 7  
|||||  
Db 681 VFFAED 686

RESULT 86

Q8GJB6

ID Q8GJB6 PRELIMINARY; PRT; 136 AA.  
AC Q8GJB6;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein.  
GN CJP39.  
OS Campylobacter jejuni.  
OG Plasmid pVir.  
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;  
OC Campylobacteraceae; Campylobacter.  
OX NCBI\_TaxID=197;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=81-176;  
RX MEDLINE=22267121; PubMed=12379703;  
RA Bacon D.J., Alm R.A., Hu L., Hickey T.E., Ewing C.P., Batchelor R.A.,  
RA Trust T.J., Guerry P.;  
RT "DNA Sequence and Mutational Analyses of the pVir Plasmid of  
RT Campylobacter jejuni 81-176."  
RL Infect. Immun. 70:6242-6250(2002).  
DR EMBL; AF226280; AAN46932.1; -.  
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
KW Hypothetical protein; Plasmid.  
SQ SEQUENCE 136 AA; 16129 MW; B2499FF56A7F8F45 CRC64;

Query Match 73.2%; Score 30; DB 2; Length 136;  
Best Local Similarity 50.0%; Pred. No. 1.9e+02;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
::||:|  
Db 20 IIFFGKDF 27

RESULT 87

Q8PEK4

ID Q8PEK4 PRELIMINARY; PRT; 137 AA.  
AC Q8PEK4;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Enterocin biosynthesis related protein.  
GN XAC4345.  
OS Xanthomonas axonopodis (pv. citri).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
OC Xanthomonadaceae; Xanthomonas.

OX NCBI\_TaxID=92829;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=306 / ATCC 13902 / XV 101;  
 RX MEDLINE=22022145; PubMed=12024217;  
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
 RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,  
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,  
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
 RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
 RA Setubal J.C., Kitajima J.P.;  
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
 RT host specificities."  
 RL Nature 417:459-463(2002).  
 DR EMBL; AE012089; AAM39175.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 137 AA; 15312 MW; 940D07B0F5EBA043 CRC64;

Query Match 73.2%; Score 30; DB 16; Length 137;  
 Best Local Similarity 57.1%; Pred. No. 1.9e+02;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 VFFAEDF 8  
 ::||:|  
 Db 37 IYFADDF 43

# RESULT 88

Q81Z55

ID Q81Z55 PRELIMINARY; PRT; 138 AA.  
 AC Q81Z55;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE General stress protein 26.  
 GN BA0418.  
 OS Bacillus anthracis (strain Ames).  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=198094;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22608414; PubMed=12721629;  
 RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,  
 RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,  
 RA Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,  
 RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,  
 RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,  
 RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,

RA Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,  
 RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,  
 RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,  
 RA Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,  
 RA Fraser C.M.;  
 RT "The genome sequence of *Bacillus anthracis* Ames and comparison to  
 RT closely related bacteria.";  
 RL Nature 423:81-86(2003).  
 DR EMBL; AE017025; AAP24447.1; -.  
 DR TIGR; BA0418; -.  
 DR InterPro; IPR009002; FMN\_binding.  
 KW Complete proteome.  
 SQ SEQUENCE 138 AA; 16068 MW; 01A47C4A37870CF5 CRC64;

Query Match 73.2%; Score 30; DB 16; Length 138;  
 Best Local Similarity 62.5%; Pred. No. 1.9e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LVFFAEDF 8  
 :||| |||  
 Db 32 MMFFHEDF 39

# RESULT 89

Q81ID0

ID Q81ID0 PRELIMINARY; PRT; 138 AA.  
 AC Q81ID0;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE General stress protein 26.  
 GN BC0461.  
 OS *Bacillus cereus* (strain ATCC 14579 / DSM 31).  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; *Bacillus*.  
 OX NCBI\_TaxID=226900;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22608415; PubMed=12721630;  
 RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,  
 RA Kapatral V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,  
 RA Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,  
 RA Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,  
 RA Overbeek R., Kyrpides N.;  
 RT "Genome sequence of *Bacillus cereus* and comparative analysis with  
 RT *Bacillus anthracis*.";  
 RL Nature 423:87-91(2003).  
 DR EMBL; AE016999; AAP07501.1; -.  
 DR InterPro; IPR009002; FMN\_binding.  
 KW Complete proteome.  
 SQ SEQUENCE 138 AA; 15967 MW; 1C0516EA5B56DD92 CRC64;

Query Match 73.2%; Score 30; DB 16; Length 138;  
 Best Local Similarity 62.5%; Pred. No. 1.9e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 :||| |||

Db

32 MMFFHEDF 39

RESULT 90

Q83U07

ID Q83U07 PRELIMINARY; PRT; 153 AA.  
AC Q83U07;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Putative mating pair formation protein (MpfA).  
GN MPFA.  
OS Pseudomonas putida.  
OG Plasmid pDTG1, and Plasmid pCg1.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=303;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCIB 9816-4; PLASMID=pDTG1;  
RA Zylstra G.J., Dennis J.J.;  
RT "Complete nucleotide sequence of the NAH plasmid pDTG1 from  
RT Pseudomonas putida NCIB 9816-4.";  
RL Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC PLASMID=pCg1;  
RA Park W., Jeon C.O., Hohnstock-Ashe A.M., Winans S.C., Zylstra G.J.,  
RA Madsen E.L.;  
RT "Identification and characterization of the conjugal transfer region  
RT of the pCg1 plasmid from naphthalene-degrading Pseudomonas pudia  
RT Cg1.";  
RL Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF491307; AAO64341.1; -.  
DR EMBL; AY249147; AAP06659.1; -.  
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
KW Plasmid.  
SQ SEQUENCE 153 AA; 16823 MW; 97DF7896B474D643 CRC64;

Query Match 73.2%; Score 30; DB 2; Length 153;  
Best Local Similarity 75.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
||||| :|  
Db 133 LVFFAPNF 140

RESULT 91

Q93RF3

ID Q93RF3 PRELIMINARY; PRT; 163 AA.  
AC Q93RF3;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Open reading frame.  
GN ORF.

OS Treponema medium.  
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.  
 OX NCBI\_TaxID=58231;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC700293;  
 RA Fukunaga M.;  
 RT "A phylogenetic analysis of a human oral spirochete Treponema medium  
 by flagellar genes.";  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AB046578; BAB62253.1; -.  
 SQ SEQUENCE 163 AA; 18137 MW; 777247A666916E0F CRC64;

Query Match 73.2%; Score 30; DB 2; Length 163;  
 Best Local Similarity 75.0%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 |||| ||  
 Db 152 LVFFTEKF 159

# RESULT 92

Q92CD6

ID Q92CD6 PRELIMINARY; PRT; 174 AA.  
 AC Q92CD6;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein lin1236.  
 GN LIN1236.  
 OS Listeria innocua.  
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
 OX NCBI\_TaxID=1642;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CLIP 11262 / Serovar 6a;  
 RX MEDLINE=21537279; PubMed=11679669;  
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
 RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,  
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
 RT "Comparative genomics of Listeria species.";  
 RL Science 294:849-852(2001).  
 DR EMBL; AL596168; CAC96467.1; -.  
 DR PIR; AC1587; AC1587.  
 DR ListiList; LIN01236; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 174 AA; 19975 MW; B18BC9ADB185EE53 CRC64;

Query Match 73.2%; Score 30; DB 16; Length 174;  
Best Local Similarity 62.5%; Pred. No. 2.4e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
|:| | | |:  
Db 145 LLFFASDY 152

RESULT 93

O64061

ID O64061 PRELIMINARY; PRT; 216 AA.  
AC O64061;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN YOMX.  
OS Bacteriophage SPBc2.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.  
OX NCBI\_TaxID=66797;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Lazarevic V., Duesterhoeft A., Soldo B., Hilbert H., Mauel C.,  
RA Karamata D.;  
RT "The complete nucleotide sequence of the Bacillus subtilis SPbetac2  
RT prophage."  
RL Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF020713; AAC13021.1; -.  
DR PIR; T12812; T12812.  
KW Hypothetical protein.  
SQ SEQUENCE 216 AA; 24179 MW; FDD0ED46D7A5DB82 CRC64;

Query Match 73.2%; Score 30; DB 9; Length 216;  
Best Local Similarity 62.5%; Pred. No. 3.1e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
| | | |:  
Db 180 LTFFGEDY 187

RESULT 94

O31960

ID O31960 PRELIMINARY; PRT; 216 AA.  
AC O31960;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE YOMX protein.  
GN YOMX.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;

RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Guiseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*.";  
 RL Nature 390:249-256(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; Z99115; CAB14037.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 216 AA; 24179 MW; FDD0ED46D7A5DB82 CRC64;

Query Match 73.2%; Score 30; DB 16; Length 216;  
 Best Local Similarity 62.5%; Pred. No. 3.1e+02;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 | || ||:  
 Db 180 LTFFGEDY 187

# RESULT 95

O16760

ID O16760 PRELIMINARY; PRT; 222 AA.

AC O16760;

DT 01-JAN-1998 (TrEMBLrel. 05, Created)

DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein.  
 GN F59E11.6.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium."  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Bradshaw H.;  
 RT "The sequence of C. elegans cosmid F59E11."  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RT "Direct Submission."  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF016685; AAG24142.1; -.  
 DR PIR; T32121; T32121.  
 DR WormPep; F59E11.6; CE11506.  
 DR InterPro; IPR002542; DUF19.  
 DR Pfam; PF01579; DUF19; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 222 AA; 25570 MW; B8DC0919A0BDA564 CRC64;

Query Match 73.2%; Score 30; DB 5; Length 222;  
 Best Local Similarity 83.3%; Pred. No. 3.1e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 FFAEDF 8  
 ||:||||  
 Db 119 FFSEDF 124

# RESULT 96

## Q8ZAM4

ID Q8ZAM4 PRELIMINARY; PRT; 258 AA.  
 AC Q8ZAM4;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Sec-independent protein translocase protein TatC (Hypothetical  
 DE protein).  
 GN TATC OR YPO3776 OR Y0454.  
 OS Yersinia pestis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Yersinia.  
 OX NCBI\_TaxID=632;



RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CO-92 / Biovar Orientalis;  
 RX MEDLINE=21470413; PubMed=11586360;  
 RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,  
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,  
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,  
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
 RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,  
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,  
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;  
 RT "Genome sequence of *Yersinia pestis*, the causative agent of plague."  
 RL Nature 413:523-527(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KIM5 / Biovar Mediaevalis;  
 RX MEDLINE=22137863; PubMed=12142430;  
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,  
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,  
 RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,  
 RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,  
 RA Perry R.D.;  
 RT "Genome sequence of *Yersinia pestis* KIM."  
 RL J. Bacteriol. 184:4601-4611(2002).  
 DR EMBL; AJ414158; CAC93243.1; -.  
 DR EMBL; AE013646; AAM84043.1; -.  
 DR PIR; AG0459; AG0459.  
 DR InterPro; IPR008277; Tat\_C.  
 DR InterPro; IPR002033; Translocase.  
 DR Pfam; PF00902; TatC; 1.  
 DR TIGRFAMs; TIGR00945; tatC; 1.  
 DR PROSITE; PS01218; TATC; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 258 AA; 28725 MW; 7B052D8306E6FDA6 CRC64;

Query Match 73.2%; Score 30; DB 16; Length 258;  
 Best Local Similarity 85.7%; Pred. No. 3.7e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
 |||||  
 Db 34 LVFFAND 40

# RESULT 97

Q8CUH9

ID Q8CUH9 PRELIMINARY; PRT; 282 AA.  
 AC Q8CUH9;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Spore coat polysaccharide synthesis (dTDP-4-dehydrorhamnose  
 DE reductase).  
 GN OB1128.  
 OS Oceanobacillus iheyensis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.  
 OX NCBI\_TaxID=182710;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HTE831 / DSM 14371 / JCM 11309;  
 RX MEDLINE=22220767; PubMed=12235376;  
 RA Takami H., Takaki Y., Uchiyama I.;  
 RT "Genome sequence of *Oceanobacillus iheyensis* isolated from the Iheya  
 RT Ridge and its unexpected adaptive capabilities to extreme  
 RT environments.";  
 RL Nucleic Acids Res. 30:3927-3935(2002).  
 DR EMBL; AP004596; BAC13084.1; -.  
 DR GO; GO:0008831; F:dTDP-4-dehydrorhamnose reductase activity; IEA.  
 DR GO; GO:0045226; P:extracellular polysaccharide biosynthesis; IEA.  
 DR InterPro; IPR005913; TDP\_rham\_reduct.  
 DR Pfam; PF04321; RmlD\_sub\_bind; 1.  
 DR TIGRFAMs; TIGR01214; rmlD; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 282 AA; 32214 MW; F53A8BE4CCD17075 CRC64;

Query Match 73.2%; Score 30; DB 16; Length 282;  
 Best Local Similarity 62.5%; Pred. No. 4e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 ||||: |:  
 Db 99 LVFFSSDY 106

# RESULT 98

Q838Z8

ID Q838Z8 PRELIMINARY; PRT; 311 AA.  
 AC Q838Z8;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Cation efflux family protein.  
 GN EF0280.  
 OS *Enterococcus faecalis* (*Streptococcus faecalis*).  
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.  
 OX NCBI\_TaxID=1351;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=V583 / ATCC 700802;  
 RX MEDLINE=22550857; PubMed=12663927;  
 RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,  
 RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,  
 RA Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,  
 RA Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,  
 RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,  
 RA Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;  
 RT "Role of mobile DNA in the evolution of vancomycin-resistant  
 RT *Enterococcus faecalis*.";  
 RL Science 299:2071-2074(2003).  
 DR EMBL; AE016947; AAO80143.1; -.  
 DR TIGR; EF0280; -.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0008324; F:cation transporter activity; IEA.  
 DR GO; GO:0006812; P:cation transport; IEA.

DR InterPro; IPR002524; Cation\_efflux.  
 DR Pfam; PF01545; Cation\_efflux; 1.  
 DR TIGRFAMs; TIGR01297; CDF; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 311 AA; 33897 MW; E2C76C9D508BA875 CRC64;

Query Match 73.2%; Score 30; DB 16; Length 311;  
 Best Local Similarity 62.5%; Pred. No. 4.4e+02;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 LVFFAEDF 8  
 | : | |||  
 Db 163 LIIFTEDF 170

# RESULT 99

Q9KE67

ID Q9KE67 PRELIMINARY; PRT; 327 AA.  
 AC Q9KE67;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE ABC transporter (Substrate-binding protein).  
 GN BH0990.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 RT halodurans and genomic sequence comparison with Bacillus subtilis."  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 DR EMBL; AP001510; BAB04709.1; -.  
 DR PIR; F83773; F83773.  
 DR GO; GO:0005381; F:iron ion transporter activity; IEA.  
 DR GO; GO:0006827; P:high affinity iron ion transport; IEA.  
 DR InterPro; IPR002491; Peripla\_BP.  
 DR InterPro; IPR000437; Prok\_lipoprot\_S.  
 DR Pfam; PF01497; Peripla\_BP\_2; 1.  
 DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 327 AA; 36362 MW; 8254A0BF90873F17 CRC64;

Query Match 73.2%; Score 30; DB 16; Length 327;  
 Best Local Similarity 71.4%; Pred. No. 4.7e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 VFFAEDF 8  
 : | ||||  
 Db 263 IFLAEDF 269

RESULT 100

Q803V8

ID Q803V8 PRELIMINARY; PRT; 343 AA.  
 AC Q803V8;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Similar to MO25 protein.  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; BC044172; AAH44172.1; -.  
 DR InterPro; IPR008938; ARM.  
 DR InterPro; IPR004892; Mo25.  
 DR Pfam; PF03204; Mo25; 1.  
 SQ SEQUENCE 343 AA; 39820 MW; F10450DA0446268A CRC64;

Query Match 73.2%; Score 30; DB 13; Length 343;  
 Best Local Similarity 62.5%; Pred. No. 4.9e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 :| |:  
 Db 160 IVLFSEDF 167

Search completed: February 28, 2004, 08:56:41  
 Job time : 112.5 secs

CC diminishes or stops. Has also significant hyperglycemic hormone  
 CC (CHH) activity.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- MASS SPECTROMETRY: MW=9193.3; MW\_ERR=0.53; METHOD=Electrospray.  
 CC -!- SIMILARITY: Belongs to the arthropod CHH/MIH/GIH/VIH hormone  
 CC family.  
 DR InterPro; IPR001166; CHH\_MIH\_GIH.  
 DR Pfam; PF01147; Crust\_neurohorm; 1.  
 DR PRINTS; PR00550; HYPRGLYCEMIC.  
 DR PROSITE; PS01250; CHH\_MIH\_GIH; 1.  
 KW Neuropeptide; Hormone.  
 FT DISULFID 7 44 BY SIMILARITY.  
 FT DISULFID 24 40 BY SIMILARITY.  
 FT DISULFID 27 53 BY SIMILARITY.  
 SQ SEQUENCE 78 AA; 9200 MW; 7285FB5FC8BA98C6 CRC64;

Query Match 65.9%; Score 27; DB 1; Length 78;  
 Best Local Similarity 83.3%; Pred. No. 54;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 FFAEDF 8  
 || |||  
 Db 45 FFNEDE 50

# RESULT 100

## MIH\_CALSI

ID MIH\_CALSI STANDARD; PRT; 113 AA.  
 AC P55321; Q23738;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Molt-inhibiting hormone precursor (MIH).  
 OS Callinectes sapidus (Blue crab).  
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
 OC Eubrachyura; Portunoidea; Portunidae; Callinectes.  
 OX NCBI\_TaxID=6763;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Eyestalk;  
 RX MEDLINE=95251700; PubMed=7537497;  
 RA Lee K.J., Elton T.S., Bej A.K., Watts S.A., Watson R.D.;  
 RT "Molecular cloning of a cDNA encoding putative molt-inhibiting  
 RT hormone from the blue crab, Callinectes sapidus."  
 RL Biochem. Biophys. Res. Commun. 209:1126-1131(1995).  
 CC -!- FUNCTION: Inhibits Y-organs where molting hormone (ecdysteroid)  
 CC is secreted. A molting cycle is initiated when MIH secretion  
 CC diminishes or stops (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: Belongs to the arthropod CHH/MIH/GIH/VIH hormone  
 CC family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U19764; AAA69029.1; -.  
DR PIR; JC4143; JC4143.  
DR InterPro; IPR001166; CHH\_MIH\_GIH.  
DR Pfam; PF01147; Crust\_neurohorm; 1.  
DR PRINTS; PR00550; HYPRGLYCEMIC.  
DR PROSITE; PS01250; CHH\_MIH\_GIH; 1.  
KW Signal; Neuropeptide; Hormone.  
FT SIGNAL 1 35 POTENTIAL.  
FT PEPTIDE 36 113 MOLT-INHIBITING HORMONE.  
FT DISULFID 42 79 BY SIMILARITY.  
FT DISULFID 59 75 BY SIMILARITY.  
FT DISULFID 62 88 BY SIMILARITY.  
SQ SEQUENCE 113 AA; 12850 MW; 86AF3AE512130A5C CRC64;

Query Match 65.9%; Score 27; DB 1; Length 113;  
Best Local Similarity 83.3%; Pred. No. 78;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 FFAEDF 8  
|| |||  
Db 80 FFNEDE 85

Search completed: February 28, 2004, 08:53:02  
Job time : 31.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 28, 2004, 08:36:39 ; Search time 60.5 Seconds  
(without alignments)  
41.721 Million cell updates/sec

Title: US-09-668-314C-84  
Perfect score: 41  
Sequence: 1 LVFFAEDE 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Match	Query	Length	DB	ID	Description
-----								

1	35	85.4	28	4	Q9UCD1	Q9ucd1 homo sapien
2	35	85.4	30	4	Q9UCA9	Q9uca9 homo sapien
3	35	85.4	33	4	Q9UC33	Q9uc33 homo sapien
4	35	85.4	79	11	O35463	O35463 cricetulus
5	35	85.4	82	4	Q16020	Q16020 homo sapien
6	35	85.4	82	4	Q16014	Q16014 homo sapien
7	35	85.4	82	4	Q16019	Q16019 homo sapien
8	35	85.4	113	13	Q8JH58	Q8jh58 chelydra se
9	35	85.4	218	11	Q8BPV5	Q8bpv5 mus musculu
10	35	85.4	295	16	Q8E547	Q8e547 streptococc
11	35	85.4	295	16	Q8DZI3	Q8dzi3 streptococc
12	35	85.4	357	13	Q8UUI8	Q8uui8 brachydanio
13	35	85.4	361	8	O20025	O20025 crithmum ma
14	35	85.4	361	8	O20011	O20011 anthriscus
15	35	85.4	364	8	O20068	O20068 neogoezia m
16	35	85.4	384	11	Q8BPC7	Q8bpc7 mus musculu
17	35	85.4	472	13	Q8UUS0	Q8uus0 brachydanio
18	35	85.4	534	13	O93296	O93296 gallus gall
19	35	85.4	569	13	Q9PVL1	Q9pvl1 gallus gall
20	35	85.4	612	13	Q9I9E7	Q9i9e7 brachydanio
21	35	85.4	678	13	Q7ZZT1	Q7zzt1 brachydanio
22	35	85.4	695	13	Q9DGJ8	Q9dgj8 gallus gall
23	35	85.4	738	13	Q90W28	Q90w28 brachydanio
24	35	85.4	751	13	Q9DGJ7	Q9dgj7 gallus gall
25	35	85.4	1169	5	Q8T9D3	Q8t9d3 drosophila
26	35	85.4	1169	5	Q9VSJ6	Q9vsj6 drosophila
27	33	80.5	222	5	Q21915	Q21915 caenorhabdi
28	33	80.5	261	6	Q9XSI7	Q9xsi7 bos taurus
29	33	80.5	448	16	Q87M09	Q87m09 vibrio para
30	33	80.5	455	17	Q50563	Q50563 methanobact
31	33	80.5	502	5	O62511	O62511 caenorhabdi
32	33	80.5	651	17	Q8THF4	Q8thf4 methanosarc
33	33	80.5	3610	5	Q968T7	Q968t7 plasmodium
34	33	80.5	3620	5	Q968T6	Q968t6 plasmodium
35	33	80.5	3628	5	Q968Y0	Q968y0 plasmodium
36	33	80.5	3704	5	Q8IKY8	Q8iky8 plasmodium
37	32	78.0	60	5	Q9BHZ8	Q9bhz8 globodera r
38	32	78.0	75	12	Q90160	Q90160 bombyx mori
39	32	78.0	184	16	Q931V3	Q931v3 staphylococ
40	32	78.0	261	2	Q7X225	Q7x225 staphylococ
41	32	78.0	261	2	Q7WRM0	Q7wrm0 staphylococ
42	32	78.0	261	16	Q99V89	Q99v89 staphylococ
43	32	78.0	268	16	Q8NXD0	Q8nxd0 staphylococ
44	32	78.0	379	16	Q83NF3	Q83nf3 tropheryma
45	32	78.0	390	16	Q83N16	Q83n16 tropheryma
46	32	78.0	398	2	O87160	O87160 vibrio chol
47	32	78.0	656	16	Q7ULM6	Q7ulm6 rhodopirell
48	32	78.0	1676	16	Q8A6R7	Q8a6r7 bacteroides
49	32	78.0	1732	4	Q7Z443	Q7z443 homo sapien
50	31	75.6	49	6	O97917	O97917 bos taurus
51	31	75.6	141	17	Q974D3	Q974d3 sulfolobus
52	31	75.6	150	5	P91460	P91460 caenorhabdi
53	31	75.6	161	16	Q8E591	Q8e591 streptococc
54	31	75.6	179	16	Q82JK4	Q82jk4 streptomyce
55	31	75.6	182	16	Q9Z588	Q9z588 streptomyce
56	31	75.6	244	5	Q9VQ40	Q9vq40 drosophila
57	31	75.6	259	16	Q8NN32	Q8nn32 corynebacte



58	31	75.6	262	6	Q9BGL3	Q9bgl3 ovis aries
59	31	75.6	291	16	Q8Y477	Q8y477 listeria mo
60	31	75.6	308	16	Q7VMQ9	Q7vmq9 haemophilus
61	31	75.6	361	8	O20015	O20015 bifora amer
62	31	75.6	368	16	O66650	O66650 aquifex aeo
63	31	75.6	370	16	Q7WD43	Q7wd43 bordetella
64	31	75.6	370	16	Q7W5K7	Q7w5k7 bordetella
65	31	75.6	370	16	Q7VW49	Q7vw49 bordetella
66	31	75.6	392	16	Q8D9D2	Q8d9d2 vibrio vuln
67	31	75.6	438	2	O06056	O06056 mycobacteri
68	31	75.6	455	17	Q8U383	Q8u383 pyrococcus
69	31	75.6	519	16	Q8R7Q5	Q8r7q5 thermoanaer
70	31	75.6	524	2	Q7X127	Q7x127 xanthomonas
71	31	75.6	524	16	Q8PRK2	Q8prk2 xanthomonas
72	31	75.6	537	4	Q96AA0	Q96aa0 homo sapien
73	31	75.6	582	6	Q9XSP0	Q9xsp0 capra hircu
74	31	75.6	582	6	Q9GLE4	Q9gle4 bos taurus
75	31	75.6	603	5	Q86C23	Q86c23 entamoeba h
76	31	75.6	693	13	Q98SG0	Q98sg0 xenopus lae
77	31	75.6	715	4	Q7Z5X1	Q7z5x1 homo sapien
78	31	75.6	747	13	Q91963	Q91963 xenopus. ap
79	31	75.6	748	5	O77066	O77066 choristoneu
80	31	75.6	937	4	Q96FT5	Q96ft5 homo sapien
81	31	75.6	956	11	Q99K64	Q99k64 mus musculu
82	31	75.6	956	11	Q8R542	Q8r542 mus musculu
83	31	75.6	1016	4	Q8NDE6	Q8nde6 homo sapien
84	31	75.6	1296	5	Q8WRY3	Q8wry3 plasmodium
85	31	75.6	1364	3	Q8NIT0	Q8nit0 neurospora
86	30	73.2	136	2	Q8GJB6	Q8gjb6 campylobact
87	30	73.2	137	16	Q8PEK4	Q8pek4 xanthomonas
88	30	73.2	138	16	Q81Z55	Q81z55 bacillus an
89	30	73.2	138	16	Q81ID0	Q8lid0 bacillus ce
90	30	73.2	153	2	Q83U07	Q83u07 pseudomonas
91	30	73.2	163	2	Q93RF3	Q93rf3 treponema m
92	30	73.2	174	16	Q92CD6	Q92cd6 listeria in
93	30	73.2	216	9	O64061	O64061 bacterioph
94	30	73.2	216	16	O31960	O31960 bacillus su
95	30	73.2	222	5	O16760	O16760 caenorhabdi
96	30	73.2	258	16	Q8ZAM4	Q8zam4 yersinia pe
97	30	73.2	282	16	Q8CUH9	Q8cuh9 oceanobacil
98	30	73.2	311	16	Q838Z8	Q838z8 enterococcu
99	30	73.2	327	16	Q9KE67	Q9ke67 bacillus ha
100	30	73.2	343	13	Q803V8	Q803v8 brachydanio
101	30	73.2	354	16	Q898W5	Q898w5 clostridium
102	30	73.2	393	17	Q8U238	Q8u238 pyrococcus
103	30	73.2	400	8	Q95BX4	Q95bx4 pteroceltis
104	30	73.2	402	16	Q98RF1	Q98rf1 mycoplasma
105	30	73.2	413	5	Q9XZR6	Q9xzr6 drosophila
106	30	73.2	441	16	Q8ZNH0	Q8znh0 salmonella
107	30	73.2	457	16	Q8ZJR8	Q8zjr8 yersinia pe
108	30	73.2	459	16	Q8CZM3	Q8czm3 yersinia pe
109	30	73.2	462	2	Q9RA87	Q9ra87 sphingomona
110	30	73.2	462	2	Q9RP80	Q9rp80 pseudomonas
111	30	73.2	471	10	Q9M394	Q9m394 arabidopsis
112	30	73.2	501	3	Q7Z857	Q7z857 phaffia rho
113	30	73.2	505	8	Q95BY1	Q95by1 celtis sine
114	30	73.2	505	8	Q85V88	Q85v88 celtis occi

115	30	73.2	527	5	Q9N4T6	Q9n4t6 caenorhabdi
116	30	73.2	566	3	Q12321	Q12321 saccharomyc
117	30	73.2	582	3	O13506	O13506 phaffia rho
118	30	73.2	582	3	Q7Z858	Q7z858 phaffia rho
119	30	73.2	590	2	Q54525	Q54525 streptococc
120	30	73.2	590	16	Q9A146	Q9a146 streptococc
121	30	73.2	590	16	Q8P270	Q8p270 streptococc
122	30	73.2	590	16	Q8K8E2	Q8k8e2 streptococc
123	30	73.2	603	16	Q878D3	Q878d3 streptococc
124	30	73.2	612	2	Q8VPV9	Q8vpv9 synechococc
125	30	73.2	664	16	Q9PN63	Q9pn63 campylobact
126	30	73.2	700	5	Q8IIJ9	Q8iij9 plasmodium
127	30	73.2	745	12	O36374	O36374 alcelaphine
128	30	73.2	755	4	O00261	O00261 homo sapien
129	30	73.2	759	4	Q96C67	Q96c67 homo sapien
130	30	73.2	768	10	Q9LFI1	Q9lfi1 arabidopsis
131	30	73.2	791	5	Q86MD8	Q86md8 halocynthia
132	30	73.2	906	16	Q8AA21	Q8aa21 bacteroides
133	30	73.2	1077	13	Q8UWA6	Q8uwa6 tribolodon
134	30	73.2	1088	13	Q8AYH7	Q8ayh7 oncorhynchu
135	30	73.2	1185	4	Q9NV86	Q9nv86 homo sapien
136	30	73.2	1265	3	O59920	O59920 pneumocysti
137	30	73.2	1797	11	Q80X19	Q80x19 mus musculu
138	30	73.2	2143	11	Q80T20	Q80t20 mus musculu
139	30	73.2	4780	5	Q8IE54	Q8ie54 plasmodium
140	29	70.7	70	12	Q84661	Q84661 paramecium
141	29	70.7	70	16	Q8PRN2	Q8prn2 xanthomonas
142	29	70.7	100	16	Q8Y8G0	Q8y8g0 listeria mo
143	29	70.7	111	16	Q9XA43	Q9xa43 streptomyce
144	29	70.7	119	6	Q9GMZ2	Q9gmz2 sus scrofa
145	29	70.7	135	5	O18088	O18088 caenorhabdi
146	29	70.7	141	2	Q8GBR7	Q8gbr7 treponema m
147	29	70.7	146	16	Q8REP1	Q8rep1 fusobacteri
148	29	70.7	167	5	Q9N4B5	Q9n4b5 caenorhabdi
149	29	70.7	168	5	Q8N0N2	Q8n0n2 littorina l
150	29	70.7	169	8	Q85WI3	Q85wi3 uncultured
151	29	70.7	173	8	Q9GGR5	Q9ggr5 prasinophyc
152	29	70.7	182	12	Q91SD3	Q91sd3 human calic
153	29	70.7	192	2	Q46756	Q46756 escherichia
154	29	70.7	193	13	Q9W7E6	Q9w7e6 oryzias lat
155	29	70.7	202	2	Q93P97	Q93p97 microscilla
156	29	70.7	206	16	Q9JQN8	Q9jqn8 neisseria m
157	29	70.7	215	10	Q94FL5	Q94fl5 arabidopsis
158	29	70.7	229	9	Q9MC82	Q9mc82 bacterioph
159	29	70.7	234	16	Q97SX9	Q97sx9 streptococc
160	29	70.7	234	16	Q8CZ91	Q8cz91 streptococc
161	29	70.7	244	17	O30150	O30150 archaeoglob
162	29	70.7	247	10	Q9FX75	Q9fx75 arabidopsis
163	29	70.7	251	17	Q9V2E8	Q9v2e8 pyrococcus
164	29	70.7	269	16	Q87RN1	Q87rn1 vibrio para
165	29	70.7	273	10	Q9T0H9	Q9t0h9 arabidopsis
166	29	70.7	281	10	Q8GU34	Q8gu34 palmaria pa
167	29	70.7	282	16	Q97TL0	Q97tl0 clostridium
168	29	70.7	306	17	Q97BE5	Q97be5 thermoplasm
169	29	70.7	310	8	O63624	O63624 dalbulus ci
170	29	70.7	310	8	O63636	O63636 baldulus tr
171	29	70.7	310	8	O63635	O63635 baldulus tr

172	29	70.7	310	8	O63631	O63631 dalbulus ma
173	29	70.7	310	8	O63633	O63633 dalbulus ma
174	29	70.7	314	16	Q9CFM1	Q9cfm1 lactococcus
175	29	70.7	332	5	P91231	P91231 caenorhabdi
176	29	70.7	333	8	O20001	O20001 aciphylla a
177	29	70.7	333	16	Q81CC4	Q81cc4 bacillus ce
178	29	70.7	334	16	Q8R8F2	Q8r8f2 thermoanaer
179	29	70.7	335	16	Q82G15	Q82g15 streptomyce
180	29	70.7	340	16	Q9KYD8	Q9kyd8 streptomyce
181	29	70.7	343	11	Q9D9P9	Q9d9p9 mus musculu
182	29	70.7	347	1	Q9P9F9	Q9p9f9 methanosarc
183	29	70.7	350	17	Q8PU59	Q8pu59 methanosarc
184	29	70.7	351	8	Q9MJ68	Q9mj68 physarum po
185	29	70.7	356	8	O20000	O20000 arracacia a
186	29	70.7	359	8	O20005	O20005 ammi majus.
187	29	70.7	360	8	O20053	O20053 heracleum l
188	29	70.7	361	8	O20058	O20058 lagoecia cu
189	29	70.7	361	8	O20009	O20009 arracacia q
190	29	70.7	361	8	O20004	O20004 angelica lu
191	29	70.7	361	8	O20089	O20089 scandix pec
192	29	70.7	361	8	O20072	O20072 petroselinu
193	29	70.7	361	8	O20003	O20003 anethum gra
194	29	70.7	361	8	O20038	O20038 donnellsmit
195	29	70.7	361	8	O20094	O20094 taenidia in
196	29	70.7	361	8	O20061	O20061 levisticum
197	29	70.7	361	8	O20081	O20081 rhodosciadi
198	29	70.7	361	8	O20033	O20033 coriandrum
199	29	70.7	361	8	O20046	O20046 foeniculum
200	29	70.7	361	8	O20084	O20084 sphenosciad
201	29	70.7	361	8	O20026	O20026 coaxana pur
202	29	70.7	361	8	O20021	O20021 cicuta doug
203	29	70.7	361	8	O20044	O20044 enantiophyl
204	29	70.7	361	8	O20085	O20085 spermolepis
205	29	70.7	363	8	O20077	O20077 pimpinella
206	29	70.7	364	8	O20088	O20088 shoshonea p
207	29	70.7	364	8	Q8SN50	Q8sn50 lilaeopsis
208	29	70.7	364	8	Q8SN48	Q8sn48 lilaeopsis
209	29	70.7	364	8	O20073	O20073 perideridia
210	29	70.7	364	8	O20090	O20090 sium suave
211	29	70.7	364	8	O20070	O20070 oenanthe sa
212	29	70.7	364	8	Q8SN49	Q8sn49 lilaeopsis
213	29	70.7	381	17	Q8TJA7	Q8tja7 methanosarc
214	29	70.7	394	5	O18098	O18098 caenorhabdi
215	29	70.7	404	13	Q90420	Q90420 brachydanio
216	29	70.7	410	10	Q7Y177	Q7y177 oryza sativ
217	29	70.7	413	2	Q8RNN1	Q8rnn1 legionella
218	29	70.7	419	6	O19134	O19134 oryctolagus
219	29	70.7	422	16	Q8A8L2	Q8a8l2 bacteroides
220	29	70.7	438	11	O35934	O35934 mus musculu
221	29	70.7	438	11	Q8C5B1	Q8c5b1 mus musculu
222	29	70.7	455	17	O58651	O58651 pyrococcus
223	29	70.7	470	13	Q9YGQ2	Q9ygq2 morone amer
224	29	70.7	488	4	Q96M73	Q96m73 homo sapien
225	29	70.7	488	16	Q8XP05	Q8xp05 clostridium
226	29	70.7	489	8	Q8M8U7	Q8m8u7 valeriana h
227	29	70.7	492	16	Q7TV33	Q7tv33 prochloroco
228	29	70.7	494	16	Q7TUC6	Q7tuc6 prochloroco

229	29	70.7	494	17	Q9HRF2	Q9hrf2 halobacteri
230	29	70.7	502	4	Q9H5C6	Q9h5c6 homo sapien
231	29	70.7	503	11	Q99LN1	Q99ln1 mus musculu
232	29	70.7	508	16	Q9RJ02	Q9rj02 streptomyce
233	29	70.7	510	8	Q9TKU3	Q9tku3 vataireopsi
234	29	70.7	513	8	O63900	O63900 myoxus glis
235	29	70.7	513	16	Q8A226	Q8a226 bacteroides
236	29	70.7	518	16	Q8XP93	Q8xp93 clostridium
237	29	70.7	519	16	Q7VPF7	Q7vpf7 haemophilus
238	29	70.7	522	16	Q895F3	Q895f3 clostridium
239	29	70.7	523	16	Q7VBM5	Q7vbm5 prochloroco
240	29	70.7	523	16	Q7U658	Q7u658 synechococc
241	29	70.7	526	13	Q7SXB7	Q7sxb7 brachydanio
242	29	70.7	528	16	Q8DBH6	Q8dbh6 vibrio vuln
243	29	70.7	528	16	Q87MC7	Q87mc7 vibrio para
244	29	70.7	534	16	Q9KTM2	Q9ktm2 vibrio chol
245	29	70.7	541	16	Q8EFD3	Q8efd3 shewanella
246	29	70.7	544	12	Q66418	Q66418 desert shie
247	29	70.7	545	12	Q918C8	Q918c8 norwalk-lik
248	29	70.7	545	12	Q9IV45	Q9iv45 human calic
249	29	70.7	545	12	Q8VA02	Q8va02 human calic
250	29	70.7	545	12	Q91H12	Q91h12 human calic
251	29	70.7	545	12	Q9WI82	Q9wi82 norwalk-lik
252	29	70.7	545	12	Q918C2	Q918c2 norwalk-lik
253	29	70.7	545	16	Q7UYV5	Q7uyv5 rhodopirell
254	29	70.7	571	12	Q03909	Q03909 influenza a
255	29	70.7	572	5	O44154	O44154 caenorhabdi
256	29	70.7	574	13	Q7T2J2	Q7t2j2 brachydanio
257	29	70.7	575	11	Q9QUU4	Q9quu4 rattus sp.
258	29	70.7	577	8	Q85QS0	Q85qs0 tricholepid
259	29	70.7	607	5	Q21921	Q21921 caenorhabdi
260	29	70.7	611	6	Q8WNQ3	Q8wnq3 sus scrofa
261	29	70.7	633	16	Q896G7	Q896g7 clostridium
262	29	70.7	659	2	Q845L0	Q845l0 bacillus me
263	29	70.7	660	16	Q8ENJ5	Q8enj5 oceanobacil
264	29	70.7	663	16	Q8E471	Q8e471 streptococc
265	29	70.7	663	16	Q8DYL6	Q8dyl6 streptococc
266	29	70.7	663	16	Q8CWX7	Q8cwx7 streptococc
267	29	70.7	693	16	Q8YQR2	Q8yqr2 anabaena sp
268	29	70.7	697	10	Q9FKE4	Q9fke4 arabidopsis
269	29	70.7	747	5	Q9Y0E5	Q9y0e5 hyphantria
270	29	70.7	750	5	Q963S9	Q963s9 plodia inte
271	29	70.7	751	5	Q9U5K5	Q9u5k5 spodoptera
272	29	70.7	753	5	Q25050	Q25050 hyphantria
273	29	70.7	753	5	O17481	O17481 hyalophora
274	29	70.7	771	11	Q80X33	Q80x33 mus musculu
275	29	70.7	816	13	Q98SV4	Q98sv4 ictalurus p
276	29	70.7	840	16	Q9RS60	Q9rs60 deinococcus
277	29	70.7	857	11	Q8K0D1	Q8k0d1 mus musculu
278	29	70.7	876	13	Q91593	Q91593 xenopus lae
279	29	70.7	898	13	Q91592	Q91592 xenopus lae
280	29	70.7	904	5	Q86AQ1	Q86aq1 dictyosteli
281	29	70.7	911	4	Q7Z7D0	Q7z7d0 homo sapien
282	29	70.7	929	11	Q80XR1	Q80xr1 mus musculu
283	29	70.7	938	11	Q8C9I6	Q8c9i6 mus musculu
284	29	70.7	957	11	Q8CI71	Q8ci71 mus musculu
285	29	70.7	964	11	Q8C6D0	Q8c6d0 mus musculu

286	29	70.7	964	11	Q7TSR5	Q7tsr5 mus musculu
287	29	70.7	1147	3	Q8X0Y3	Q8x0y3 neurospora
288	29	70.7	1149	10	Q7XN28	Q7xn28 oryza sativ
289	29	70.7	1292	5	Q95NI9	Q95ni9 plasmodium
290	29	70.7	1340	10	Q9LTT9	Q9ltt9 arabidopsis
291	29	70.7	1344	10	Q8L7N9	Q8l7n9 arabidopsis
292	29	70.7	1344	10	Q9LTT8	Q9ltt8 arabidopsis
293	29	70.7	1749	5	Q86IN4	Q86in4 dictyosteli
294	29	70.7	2151	11	Q7TN87	Q7tn87 mus musculu
295	29	70.7	2391	10	Q7XR11	Q7xr11 oryza sativ
296	29	70.7	2470	5	Q7YYJ2	Q7yyj2 cryptospori
297	29	70.7	2581	5	Q9N4C8	Q9n4c8 caenorhabdi
298	28.5	69.5	639	2	Q52781	Q52781 rhizobium l
299	28.5	69.5	639	2	O08228	O08228 rhizobium l
300	28	68.3	36	16	Q8EY51	Q8ey51 leptospira
301	28	68.3	47	16	Q8EZN7	Q8ezn7 leptospira
302	28	68.3	51	9	Q7Y3H8	Q7y3h8 enterobacte
303	28	68.3	98	16	Q8DJ04	Q8dj04 synechococc
304	28	68.3	105	5	O01304	O01304 armadillidi
305	28	68.3	107	16	Q8RA36	Q8ra36 thermoanaer
306	28	68.3	109	12	Q9DR78	Q9dr78 bovine coro
307	28	68.3	109	12	Q84732	Q84732 porcine hem
308	28	68.3	109	12	Q9QAQ5	Q9qaq5 bovine coro
309	28	68.3	109	12	Q8JSP6	Q8jsp6 porcine hem
310	28	68.3	109	12	Q99H66	Q99h66 bovine coro
311	28	68.3	109	12	Q8V434	Q8v434 bovine coro
312	28	68.3	109	12	Q9PY31	Q9py31 bovine coro
313	28	68.3	109	12	Q89792	Q89792 porcine hem
314	28	68.3	109	12	Q9QAR2	Q9qar2 bovine coro
315	28	68.3	127	12	Q9WH22	Q9wh22 indian mung
316	28	68.3	127	12	Q913F1	Q913f1 indian mung
317	28	68.3	134	12	Q91N72	Q91n72 mungbean ye
318	28	68.3	134	12	Q917N3	Q917n3 indian mung
319	28	68.3	134	12	Q910Y9	Q910y9 soybean yel
320	28	68.3	134	12	Q8JKF5	Q8j kf5 cowpea gold
321	28	68.3	134	12	Q80P65	Q80p65 legume yell
322	28	68.3	134	12	Q80P60	Q80p60 legume yell
323	28	68.3	134	12	Q80P49	Q80p49 legume yell
324	28	68.3	134	12	Q808V1	Q808v1 mungbean ye
325	28	68.3	134	12	Q808U3	Q808u3 mungbean ye
326	28	68.3	134	12	Q80P54	Q80p54 legume yell
327	28	68.3	134	12	Q7TLE4	Q7tle4 legume yell
328	28	68.3	135	10	Q8LLV9	Q8llv9 chlamydomon
329	28	68.3	140	10	Q9FNY8	Q9fny8 arabidopsis
330	28	68.3	141	16	Q9I3Z0	Q9i3z0 pseudomonas
331	28	68.3	141	17	Q9HN49	Q9hn49 halobacteri
332	28	68.3	142	16	Q7UG51	Q7ug51 rhodopirell
333	28	68.3	155	16	Q8P4C0	Q8p4c0 xanthomonas
334	28	68.3	161	13	Q9DFW6	Q9dfw6 tylosurus c
335	28	68.3	162	8	Q32406	Q32406 heteranther
336	28	68.3	175	16	Q8ES22	Q8es22 oceanobacil
337	28	68.3	179	5	Q8I2K3	Q8i2k3 plasmodium
338	28	68.3	183	12	Q98243	Q98243 molluscum c
339	28	68.3	186	11	Q8K255	Q8k255 mus musculu
340	28	68.3	186	16	Q8EPV6	Q8epv6 oceanobacil
341	28	68.3	187	10	Q7XW72	Q7xw72 oryza sativ
342	28	68.3	189	5	O01648	O01648 drosophila

343	28	68.3	189	10	Q8L6Z0	Q8l6z0 arabidopsis
344	28	68.3	190	5	O62320	O62320 caenorhabdi
345	28	68.3	190	16	Q92GV4	Q92gv4 rickettsia
346	28	68.3	193	16	Q8XIP8	Q8xip8 clostridium
347	28	68.3	195	16	Q9L0K7	Q9l0k7 streptomyce
348	28	68.3	197	10	Q94F76	Q94f76 zea mays (m
349	28	68.3	197	16	Q7VR77	Q7vr77 candidatus
350	28	68.3	213	4	Q9UJG1	Q9ujg1 homo sapien
351	28	68.3	213	11	Q8VEL0	Q8vel0 mus musculu
352	28	68.3	214	11	Q9D8Y9	Q9d8y9 mus musculu
353	28	68.3	215	10	Q7XQR5	Q7xqr5 oryza sativ
354	28	68.3	215	13	Q7SYQ6	Q7syq6 xenopus lae
355	28	68.3	218	11	Q8BFS7	Q8bfs7 mus musculu
356	28	68.3	221	10	Q9FLC1	Q9flc1 arabidopsis
357	28	68.3	224	16	Q9L2I3	Q9l2i3 streptomyce
358	28	68.3	225	10	Q8VZA9	Q8vza9 arabidopsis
359	28	68.3	225	10	O82259	O82259 arabidopsis
360	28	68.3	238	16	Q82BC5	Q82bc5 streptomyce
361	28	68.3	241	16	O51573	O51573 borrelia bu
362	28	68.3	245	16	Q8NSW8	Q8nsw8 corynebacte
363	28	68.3	253	17	O57868	O57868 pyrococcus
364	28	68.3	256	16	Q98RI4	Q98ri4 mycoplasma
365	28	68.3	259	5	Q23406	Q23406 caenorhabdi
366	28	68.3	260	3	Q9UUB8	Q9uub8 schizosacch
367	28	68.3	263	10	O80476	O80476 arabidopsis
368	28	68.3	263	10	Q8LCI0	Q8lci0 arabidopsis
369	28	68.3	264	16	Q81XQ5	Q81xq5 bacillus an
370	28	68.3	265	5	Q8IE58	Q8ie58 plasmodium
371	28	68.3	269	3	Q9P3V9	Q9p3v9 schizosacch
372	28	68.3	272	10	Q94GV5	Q94gv5 oryza sativ
373	28	68.3	280	3	Q9HGQ1	Q9hgq1 schizosacch
374	28	68.3	280	3	Q9P3E5	Q9p3e5 schizosacch
375	28	68.3	281	16	Q83EA8	Q83ea8 coxiella bu
376	28	68.3	283	10	Q9LIG6	Q9lig6 arabidopsis
377	28	68.3	291	16	Q927R6	Q927r6 listeria in
378	28	68.3	300	16	Q929N0	Q929n0 listeria in
379	28	68.3	300	16	Q8Y5C6	Q8y5c6 listeria mo
380	28	68.3	303	16	Q7VI73	Q7vi73 helicobacte
381	28	68.3	310	8	O63625	O63625 dalbulus eb
382	28	68.3	316	16	Q8R9G3	Q8r9g3 thermoanaer
383	28	68.3	320	5	Q20580	Q20580 caenorhabdi
384	28	68.3	326	16	Q8FQ79	Q8fq79 corynebacte
385	28	68.3	332	10	Q9SN75	Q9sn75 arabidopsis
386	28	68.3	333	4	Q96A95	Q96a95 homo sapien
387	28	68.3	340	2	Q9AGN6	Q9agn6 clostridium
388	28	68.3	340	6	Q8HZ90	Q8hz90 saguinus oe
389	28	68.3	340	16	Q932I6	Q932i6 staphylococ
390	28	68.3	344	16	Q8UFI5	Q8ufi5 agrobacteri
391	28	68.3	345	8	Q9GI32	Q9gi32 penthorum s
392	28	68.3	348	16	O67217	O67217 aquifex aeo
393	28	68.3	350	16	Q8F131	Q8f131 leptospira
394	28	68.3	353	13	Q8AVX5	Q8avx5 xenopus lae
395	28	68.3	361	8	O20064	O20064 myrrhis odo
396	28	68.3	364	2	Q93CU7	Q93cu7 shigella bo
397	28	68.3	364	11	Q9EQL6	Q9eq16 rattus norv
398	28	68.3	365	16	Q8YVY6	Q8yvy6 anabaena sp
399	28	68.3	369	4	Q9UPZ2	Q9upz2 homo sapien

400	28	68.3	369	5	Q9Y021	Q9y021 trypanosoma
401	28	68.3	372	4	Q9H2Y8	Q9h2y8 homo sapien
402	28	68.3	372	11	Q8K4K4	Q8k4k4 mus musculu
403	28	68.3	372	11	Q91W04	Q91w04 mus musculu
404	28	68.3	373	10	Q84VE8	Q84ve8 oryza sativ
405	28	68.3	374	16	Q87PN8	Q87pn8 vibrio para
406	28	68.3	376	16	Q8XZU7	Q8xzu7 ralstonia s
407	28	68.3	379	11	Q8BJR9	Q8bjr9 mus musculu
408	28	68.3	384	8	Q8HDG6	Q8hdg6 pteridomona
409	28	68.3	384	16	O83402	O83402 treponema p
410	28	68.3	387	8	Q7YMT9	Q7ymt9 dictyocha f
411	28	68.3	388	11	Q8BZX3	Q8bzx3 mus musculu
412	28	68.3	400	16	Q8DKI0	Q8dki0 synechococc
413	28	68.3	405	8	Q8MAA4	Q8maa4 rhodella vi
414	28	68.3	415	5	Q8WRU9	Q8wru9 meloidogyne
415	28	68.3	419	2	Q9X4U3	Q9x4u3 streptococc
416	28	68.3	420	10	Q9C6D2	Q9c6d2 arabidopsis
417	28	68.3	422	16	Q97SC5	Q97sc5 streptococc
418	28	68.3	423	4	Q8IUP7	Q8iup7 homo sapien
419	28	68.3	424	2	Q9KGQ9	Q9kgq9 vibrio para
420	28	68.3	426	10	Q84QA8	Q84qa8 oryza sativ
421	28	68.3	427	10	Q7XNT0	Q7xnt0 oryza sativ
422	28	68.3	429	2	Q83VE6	Q83ve6 lactococcus
423	28	68.3	432	16	Q814I9	Q814i9 bacillus ce
424	28	68.3	433	16	Q81JL1	Q81jl1 bacillus an
425	28	68.3	433	17	Q97U98	Q97u98 sulfolobus
426	28	68.3	447	12	Q64892	Q64892 mouse adeno
427	28	68.3	447	16	Q82U80	Q82u80 nitrosomona
428	28	68.3	449	16	Q81HX4	Q81hx4 bacillus ce
429	28	68.3	457	8	Q8HDG7	Q8hdg7 pedinella s
430	28	68.3	466	10	O80785	O80785 arabidopsis
431	28	68.3	466	16	Q8D5X3	Q8d5x3 vibrio vuln
432	28	68.3	470	2	Q55999	Q55999 sphingomona
433	28	68.3	472	13	Q10833	Q10833 xenopus lae
434	28	68.3	482	10	Q8GX97	Q8gx97 arabidopsis
435	28	68.3	486	2	Q9RFC6	Q9rfc6 rhodobacter
436	28	68.3	487	2	O07342	O07342 streptococc
437	28	68.3	489	17	Q8TM03	Q8tm03 methanosarc
438	28	68.3	490	16	O83360	O83360 treponema p
439	28	68.3	501	8	Q9TLE2	Q9tle2 gaultheria
440	28	68.3	502	5	Q9GSE6	Q9gse6 tenebrio mo
441	28	68.3	504	9	Q9MBW8	Q9mbw8 bacterioph
442	28	68.3	504	16	Q87PF0	Q87pf0 vibrio para
443	28	68.3	509	10	Q9C8J7	Q9c8j7 arabidopsis
444	28	68.3	513	16	Q8EMQ1	Q8emq1 oceanobacil
445	28	68.3	516	2	Q8VMG2	Q8vmg2 pseudomonas
446	28	68.3	518	16	Q9RKZ7	Q9rkz7 streptomyce
447	28	68.3	523	2	P95618	P95618 rhodocyclus
448	28	68.3	525	2	Q9JPB5	Q9jpb5 rhodocyclus
449	28	68.3	529	10	Q9FMV0	Q9fmv0 arabidopsis
450	28	68.3	535	16	Q926G8	Q926g8 rhizobium m
451	28	68.3	537	2	P96429	P96429 rhizobium m
452	28	68.3	541	16	Q7WPA4	Q7wpa4 bordetella
453	28	68.3	551	16	Q7UDT0	Q7udt0 shigella fl
454	28	68.3	552	16	Q8XA02	Q8xa02 escherichia
455	28	68.3	554	2	Q8RTF6	Q8rtf6 oenococcus
456	28	68.3	554	10	Q9M150	Q9m150 arabidopsis

457	28	68.3	555	2	Q8RTF7	Q8rtf7 oenococcus
458	28	68.3	555	16	Q97E08	Q97e08 clostridium
459	28	68.3	558	16	Q8FL80	Q8fl80 escherichia
460	28	68.3	558	16	Q83SP3	Q83sp3 shigella fl
461	28	68.3	559	10	O04614	O04614 arabidopsis
462	28	68.3	559	16	O83432	O83432 treponema p
463	28	68.3	561	10	Q9LTJ0	Q9ltj0 arabidopsis
464	28	68.3	564	16	Q89R39	Q89r39 bradyrhizob
465	28	68.3	574	16	Q8CVD8	Q8cvd8 leptospira
466	28	68.3	590	8	Q34192	Q34192 crithidia o
467	28	68.3	590	10	Q9STC7	Q9stc7 dunaliella
468	28	68.3	594	17	Q8TIC1	Q8tic1 methanosarc
469	28	68.3	600	9	Q8LTT3	Q8ltt3 bacterioph
470	28	68.3	600	9	Q9AZC9	Q9azc9 bacterioph
471	28	68.3	600	16	Q8YJX0	Q8yjsx0 anabaena sp
472	28	68.3	623	16	Q9PJ67	Q9pj67 campylobact
473	28	68.3	628	5	Q86N93	Q86n93 nephila ina
474	28	68.3	631	16	Q7UZQ2	Q7uzq2 prochloroco
475	28	68.3	634	16	Q7V9R2	Q7v9r2 prochloroco
476	28	68.3	639	16	Q8XP76	Q8xp76 clostridium
477	28	68.3	640	16	Q8XKR0	Q8xkr0 clostridium
478	28	68.3	646	4	Q8NEJ2	Q8nej2 homo sapien
479	28	68.3	662	16	Q8D762	Q8d762 vibrio vuln
480	28	68.3	670	4	Q9Y6R3	Q9y6r3 homo sapien
481	28	68.3	695	13	Q98SF9	Q98sf9 xenopus lae
482	28	68.3	695	13	Q7ZXQ0	Q7zxq0 xenopus lae
483	28	68.3	698	11	Q8QZR9	Q8qzr9 mus musculu
484	28	68.3	701	16	Q89XI8	Q89xi8 bradyrhizob
485	28	68.3	703	5	Q9GQ56	Q9gq56 corcyra cep
486	28	68.3	704	11	Q9DA98	Q9da98 mus musculu
487	28	68.3	709	11	Q99P25	Q99p25 mus musculu
488	28	68.3	721	12	Q9YYQ1	Q9yyq1 avian adeno
489	28	68.3	728	10	Q8S0A6	Q8s0a6 oryza sativ
490	28	68.3	752	5	Q25518	Q25518 manduca sex
491	28	68.3	752	5	Q25517	Q25517 manduca sex
492	28	68.3	752	5	Q25516	Q25516 manduca sex
493	28	68.3	757	5	Q960E5	Q960e5 drosophila
494	28	68.3	760	16	Q89XI7	Q89xi7 bradyrhizob
495	28	68.3	761	5	Q9W557	Q9w557 drosophila
496	28	68.3	765	5	Q963T0	Q963t0 plodia inte
497	28	68.3	876	16	Q8XIV2	Q8xiv2 clostridium
498	28	68.3	934	16	Q88WB1	Q88wb1 lactobacill
499	28	68.3	995	4	Q9NRZ1	Q9nrz1 homo sapien
500	28	68.3	1020	2	Q9F7M0	Q9f7m0 gamma-prote
501	28	68.3	1027	16	Q8Z028	Q8z028 anabaena sp
502	28	68.3	1035	4	O15153	O15153 homo sapien
503	28	68.3	1035	6	O97915	O97915 oryctolagus
504	28	68.3	1035	11	Q9R1C4	Q9rlc4 mus musculu
505	28	68.3	1035	11	O35422	O35422 rattus norv
506	28	68.3	1035	11	O54815	O54815 rattus norv
507	28	68.3	1035	13	O13134	O13134 ambystoma t
508	28	68.3	1079	4	Q9Y6R1	Q9y6r1 homo sapien
509	28	68.3	1079	4	Q9UIC0	Q9uic0 homo sapien
510	28	68.3	1079	4	Q9UIC1	Q9uic1 homo sapien
511	28	68.3	1079	4	Q9H262	Q9h262 homo sapien
512	28	68.3	1079	4	Q9UP50	Q9up50 homo sapien
513	28	68.3	1079	6	Q9GL77	Q9gl77 bos taurus



514	28	68.3	1079	6	Q9XSZ4	Q9xsz4 oryctolagus
515	28	68.3	1079	11	Q9QXH6	Q9qxh6 rattus norv
516	28	68.3	1079	11	O88343	O88343 mus musculu
517	28	68.3	1079	11	Q9JI66	Q9ji66 rattus norv
518	28	68.3	1079	11	Q9QYA6	Q9qya6 rattus norv
519	28	68.3	1088	10	Q9C9D7	Q9c9d7 arabidopsis
520	28	68.3	1094	11	Q9JJ32	Q9jj32 rattus norv
521	28	68.3	1146	5	Q8WSX5	Q8wsx5 paramecium
522	28	68.3	1150	5	Q9W0H4	Q9w0h4 drosophila
523	28	68.3	1160	5	Q27829	Q27829 paramecium
524	28	68.3	1200	5	Q8T5I7	Q8t5i7 anopheles g
525	28	68.3	1251	5	O16568	O16568 caenorhabdi
526	28	68.3	1405	5	Q9N694	Q9n694 toxoplasma
527	28	68.3	1450	11	O54728	O54728 rattus norv
528	28	68.3	1653	2	Q9LBG3	Q9lbq3 streptococc
529	28	68.3	1721	5	Q9UAR3	Q9uar3 pacifastacu
530	28	68.3	2400	5	Q8I3U7	Q8i3u7 plasmodium
531	28	68.3	4199	16	P74440	P74440 synechocyst
532	28	68.3	5687	5	Q8IJH4	Q8ijh4 plasmodium
533	28	68.3	5910	5	Q8IAP1	Q8iap1 plasmodium
534	27.5	67.1	516	16	Q9K8P6	Q9k8p6 bacillus ha
535	27.5	67.1	516	16	Q8RAJ4	Q8raj4 thermoanaer
536	27	65.9	36	16	Q9KLP7	Q9klp7 vibrio chol
537	27	65.9	53	9	O80171	O80171 streptococc
538	27	65.9	58	17	Q8ZV06	Q8zv06 pyrobaculum
539	27	65.9	61	16	Q9PEN8	Q9pen8 xylella fas
540	27	65.9	61	16	Q8EJA5	Q8eja5 shewanella
541	27	65.9	61	16	Q81GQ4	Q81gq4 bacillus ce
542	27	65.9	75	16	O51426	O51426 borrelia bu
543	27	65.9	79	16	Q8A8P7	Q8a8p7 bacteroides
544	27	65.9	79	16	Q8A554	Q8a554 bacteroides
545	27	65.9	81	16	Q9X0Z5	Q9x0z5 thermotoga
546	27	65.9	82	16	Q8KB13	Q8kb13 chlorobium
547	27	65.9	88	16	Q7UR07	Q7ur07 rhodopirell
548	27	65.9	91	12	Q9DUB2	Q9dub2 tt virus. o
549	27	65.9	92	16	Q9HWW8	Q9hww8 pseudomonas
550	27	65.9	95	16	Q8ZA71	Q8za71 yersinia pe
551	27	65.9	100	7	O98034	O98034 homo sapien
552	27	65.9	101	10	Q9LIY8	Q9liy8 oryza sativ
553	27	65.9	104	5	Q8T4J5	Q8t4j5 scylla serr
554	27	65.9	104	16	Q8YTV9	Q8ytv9 anabaena sp
555	27	65.9	105	9	Q7Y598	Q7y598 bacterioph
556	27	65.9	111	16	Q9ACQ0	Q9acq0 streptomyce
557	27	65.9	113	5	Q9GYE8	Q9gye8 cancer pagu
558	27	65.9	119	16	Q8D9V8	Q8d9v8 vibrio vuln
559	27	65.9	120	10	Q8L4B5	Q8l4b5 oryza sativ
560	27	65.9	122	2	Q7WX48	Q7wx48 alcaligenes
561	27	65.9	122	6	O77630	O77630 bos taurus
562	27	65.9	128	16	Q9RST6	Q9rst6 deinococcus
563	27	65.9	128	16	Q9I430	Q9i430 pseudomonas
564	27	65.9	133	16	Q9ADJ5	Q9adj5 streptomyce
565	27	65.9	136	16	Q987H0	Q987h0 rhizobium l
566	27	65.9	137	16	Q8Y6R3	Q8y6r3 listeria mo
567	27	65.9	137	16	Q87LG2	Q87lg2 vibrio para
568	27	65.9	140	9	Q8SC78	Q8sc78 stx2 conver
569	27	65.9	140	9	Q7Y384	Q7y384 stx1 conver
570	27	65.9	140	9	Q7Y2S3	Q7y2s3 stx2 conver

571	27	65.9	140	16	Q9KUU1	Q9kuu1 vibrio chol
572	27	65.9	140	16	Q8DCF3	Q8dcf3 vibrio vuln
573	27	65.9	142	2	Q9RII2	Q9rii2 streptococc
574	27	65.9	143	16	Q8XS44	Q8xs44 ralstonia s
575	27	65.9	143	16	Q7WNA3	Q7wna3 bordetella
576	27	65.9	143	16	Q7W0X9	Q7w0x9 bordetella
577	27	65.9	146	16	Q99XW3	Q99xw3 streptococc
578	27	65.9	146	16	Q8NZ92	Q8nz92 streptococc
579	27	65.9	148	5	Q9VUG5	Q9vug5 drosophila
580	27	65.9	149	17	O51955	O51955 halobacteri
581	27	65.9	150	16	Q9RVS4	Q9rvs4 deinococcus
582	27	65.9	153	16	Q99TR0	Q99tr0 staphylococ
583	27	65.9	154	16	Q831M9	Q831m9 enterococcu
584	27	65.9	155	5	Q8I0Z4	Q8i0z4 caenorhabdi
585	27	65.9	158	2	Q49737	Q49737 mycobacteri
586	27	65.9	158	16	Q98LD9	Q98ld9 rhizobium l
587	27	65.9	159	13	Q8AYG7	Q8ayg7 gallus gall
588	27	65.9	160	16	Q99VY8	Q99vy8 staphylococ
589	27	65.9	160	16	Q8NXT0	Q8nxt0 staphylococ
590	27	65.9	161	2	O30617	O30617 myxococcus
591	27	65.9	164	2	Q8GMX4	Q8gmx4 methanotrop
592	27	65.9	164	11	Q8BHP4	Q8bhp4 mus musculu
593	27	65.9	167	16	Q9PML2	Q9pml2 campylobact
594	27	65.9	176	2	Q8RT79	Q8rt79 listeria mo
595	27	65.9	186	16	Q55539	Q55539 synechocyst
596	27	65.9	186	16	Q92BN0	Q92bn0 listeria in
597	27	65.9	186	16	Q8Y739	Q8y739 listeria mo
598	27	65.9	187	16	Q81SY7	Q81sy7 bacillus an
599	27	65.9	187	16	Q81FT3	Q81ft3 bacillus ce
600	27	65.9	188	16	Q8F442	Q8f442 leptospira
601	27	65.9	189	5	Q8IP62	Q8ip62 drosophila
602	27	65.9	190	17	Q9HSH9	Q9hsh9 halobacteri
603	27	65.9	193	16	Q8YL39	Q8yl39 anabaena sp
604	27	65.9	196	5	Q9VUM4	Q9vum4 drosophila
605	27	65.9	198	2	Q8GKA2	Q8gka2 helicobacte
606	27	65.9	200	11	Q8CF80	Q8cf80 mus musculu
607	27	65.9	203	4	Q9H3I4	Q9h3i4 homo sapien
608	27	65.9	204	16	Q97IZ2	Q97iz2 clostridium
609	27	65.9	205	16	Q81SM2	Q81sm2 bacillus an
610	27	65.9	215	16	Q87IG4	Q87ig4 vibrio para
611	27	65.9	216	11	Q8BHN8	Q8bhn8 mus musculu
612	27	65.9	218	16	Q89W59	Q89w59 bradyrhizob
613	27	65.9	219	16	Q81FI4	Q81fi4 bacillus ce
614	27	65.9	221	17	O58889	O58889 pyrococcus
615	27	65.9	222	16	Q8PK19	Q8pk19 xanthomonas
616	27	65.9	224	12	Q9Q8L0	Q9q8l0 myxoma viru
617	27	65.9	224	12	Q9Q8Y8	Q9q8y8 shope fibro
618	27	65.9	224	16	Q9AC07	Q9ac07 caulobacter
619	27	65.9	225	9	Q7Y4L0	Q7y4l0 streptococc
620	27	65.9	228	16	Q97KT2	Q97kt2 clostridium
621	27	65.9	230	2	Q7X0C5	Q7x0c5 gamma-prote
622	27	65.9	230	4	Q7Z5B1	Q7z5b1 homo sapien
623	27	65.9	230	16	Q9PHW2	Q9phw2 campylobact
624	27	65.9	233	5	O44644	O44644 caenorhabdi
625	27	65.9	234	16	Q99YB2	Q99yb2 streptococc
626	27	65.9	235	1	P71164	P71164 desulfuroco
627	27	65.9	236	2	Q9FB08	Q9fb08 mycoplasma

628	27	65.9	236	4	Q8WUD3	Q8wud3 homo sapien
629	27	65.9	236	11	Q8BMU2	Q8bmu2 mus musculu
630	27	65.9	236	11	Q9D4V7	Q9d4v7 mus musculu
631	27	65.9	239	3	Q92387	Q92387 trigonopsis
632	27	65.9	239	16	Q9RL46	Q9rl46 streptomyce
633	27	65.9	240	3	O93877	O93877 fusarium ox
634	27	65.9	240	16	Q822E0	Q822e0 chlamydophi
635	27	65.9	242	3	Q04701	Q04701 fusarium so
636	27	65.9	242	4	Q7Z2Y4	Q7z2y4 homo sapien
637	27	65.9	245	16	Q9CKD2	Q9ckd2 pasteurilla
638	27	65.9	246	16	Q83L46	Q83l46 shigella fl
639	27	65.9	248	16	Q8ZG63	Q8zg63 yersinia pe
640	27	65.9	249	16	Q8YL29	Q8yl29 anabaena sp
641	27	65.9	253	16	Q83B98	Q83b98 coxiella bu
642	27	65.9	254	4	Q7Z5B2	Q7z5b2 homo sapien
643	27	65.9	255	10	Q8LFD3	Q8lfd3 arabidopsis
644	27	65.9	258	10	Q84SJ7	Q84sj7 oryza sativ
645	27	65.9	258	16	Q8REW6	Q8rew6 fusobacteri
646	27	65.9	258	16	Q8FBI6	Q8fbi6 escherichia
647	27	65.9	258	16	Q8CT67	Q8ct67 staphylococ
648	27	65.9	258	16	Q83PG7	Q83pg7 shigella fl
649	27	65.9	260	10	Q7XHJ8	Q7xhj8 lycopersico
650	27	65.9	264	16	Q97IX6	Q97ix6 clostridium
651	27	65.9	265	16	Q9Z9J1	Q9z9j1 bacillus ha
652	27	65.9	266	5	Q8SV83	Q8sv83 encephalito
653	27	65.9	266	5	Q8ST65	Q8st65 encephalito
654	27	65.9	266	16	Q8A3T0	Q8a3t0 bacteroides
655	27	65.9	267	5	Q8SVF5	Q8svf5 encephalito
656	27	65.9	270	16	P74439	P74439 synechocyst
657	27	65.9	271	5	Q93319	Q93319 caenorhabdi
658	27	65.9	272	4	Q7Z2S1	Q7z2s1 homo sapien
659	27	65.9	273	16	Q8YMB8	Q8ymb8 anabaena sp
660	27	65.9	275	16	Q8F4P8	Q8f4p8 leptospira
661	27	65.9	276	5	O44646	O44646 caenorhabdi
662	27	65.9	276	16	Q8F0K4	Q8f0k4 leptospira
663	27	65.9	276	16	Q8EUB1	Q8eub1 mycoplasma
664	27	65.9	281	16	Q8EUI6	Q8eui6 mycoplasma
665	27	65.9	281	17	O29061	O29061 archaeoglob
666	27	65.9	282	10	Q9ZVG2	Q9zvg2 arabidopsis
667	27	65.9	283	5	Q8SY96	Q8sy96 drosophila
668	27	65.9	286	5	Q9U4F4	Q9u4f4 drosophila
669	27	65.9	292	17	Q9Y9R9	Q9y9r9 aeropyrum p
670	27	65.9	293	4	Q8IZ81	Q8iz81 homo sapien
671	27	65.9	293	10	Q9LXR3	Q9lxr3 arabidopsis
672	27	65.9	293	11	Q8BHP5	Q8bhp5 mus musculu
673	27	65.9	293	11	Q8BGF6	Q8bgf6 mus musculu
674	27	65.9	298	16	Q8CUH2	Q8cuh2 oceanobacil
675	27	65.9	299	17	Q97BC0	Q97bc0 thermoplasm
676	27	65.9	300	2	Q9AEC8	Q9aec8 lactobacill
677	27	65.9	300	16	Q7V283	Q7v283 prochloroco
678	27	65.9	308	10	Q8LD18	Q8ldl8 arabidopsis
679	27	65.9	309	16	Q7UV34	Q7uv34 rhodopirell
680	27	65.9	310	8	O63638	O63638 macrosteles
681	27	65.9	310	8	O63630	O63630 dalbulus lo
682	27	65.9	310	8	O63629	O63629 dalbulus gu
683	27	65.9	311	13	Q802Z6	Q802z6 brachydanio
684	27	65.9	314	16	Q9K834	Q9k834 bacillus ha

685	27	65.9	315	5	Q7YYR2	Q7yyr2 cryptospori
686	27	65.9	315	16	Q99S53	Q99s53 staphylococ
687	27	65.9	317	2	Q8RL22	Q8rl22 providencia
688	27	65.9	317	13	Q7T180	Q7t180 gallus gall
689	27	65.9	318	17	Q96Z66	Q96z66 sulfolobus
690	27	65.9	319	5	Q8STM2	Q8stm2 encephalito
691	27	65.9	320	17	Q97AV8	Q97av8 thermoplasm
692	27	65.9	323	5	O45475	O45475 caenorhabdi
693	27	65.9	323	5	O62050	O62050 caenorhabdi
694	27	65.9	323	16	Q9CHJ5	Q9chj5 lactococcus
695	27	65.9	324	16	Q8EAY2	Q8eay2 shewanella
696	27	65.9	326	2	Q44403	Q44403 agrobacteri
697	27	65.9	326	16	Q898Z4	Q898z4 clostridium
698	27	65.9	328	10	Q9FYF3	Q9fyf3 arabidopsis
699	27	65.9	333	16	Q7WMC2	Q7wmc2 bordetella
700	27	65.9	337	10	Q8H110	Q8h110 arabidopsis
701	27	65.9	337	16	Q93N67	Q93n67 coxiella bu
702	27	65.9	339	10	Q9SFZ7	Q9sfz7 arabidopsis
703	27	65.9	340	5	Q869P7	Q869p7 dictyosteli
704	27	65.9	345	16	Q8XPV7	Q8xpv7 ralstonia s
705	27	65.9	347	16	Q92XW3	Q92xw3 rhizobium m
706	27	65.9	350	11	Q923H4	Q923h4 mus musculu
707	27	65.9	358	8	Q9TLM2	Q9tlm2 compsopogon
708	27	65.9	360	2	Q9KI41	Q9ki41 agrobacteri
709	27	65.9	360	2	Q9ADZ7	Q9adz7 agrobacteri
710	27	65.9	361	16	Q92TH6	Q92th6 rhizobium m
711	27	65.9	363	5	O17784	O17784 caenorhabdi
712	27	65.9	365	12	Q91TQ9	Q91tq9 tupaia herp
713	27	65.9	366	10	Q9M7D5	Q9m7d5 ipomoea bat
714	27	65.9	366	10	Q9AUC5	Q9auc5 ipomoea bat
715	27	65.9	368	3	Q9UT60	Q9ut60 schizosacch
716	27	65.9	368	10	Q9M7D6	Q9m7d6 ipomoea bat
717	27	65.9	368	10	Q84RM9	Q84rm9 ipomoea bat
718	27	65.9	368	10	Q84RM8	Q84rm8 ipomoea bat
719	27	65.9	372	8	Q9THP4	Q9thp4 erythroclad
720	27	65.9	372	8	Q9TLM6	Q9tlm6 compsopogon
721	27	65.9	372	8	Q9TLM3	Q9tlm3 erythrotic
722	27	65.9	372	16	Q8EPT3	Q8ept3 oceanobacil
723	27	65.9	373	8	Q8WI59	Q8wi59 compsopogon
724	27	65.9	374	5	Q9TXY9	Q9txy9 caenorhabdi
725	27	65.9	374	16	Q927X3	Q927x3 listeria in
726	27	65.9	375	8	Q9TLM4	Q9tlm4 compsopogon
727	27	65.9	375	16	Q88BQ5	Q88bq5 pseudomonas
728	27	65.9	376	16	Q97LY4	Q97ly4 clostridium
729	27	65.9	380	8	Q9TLM5	Q9tlm5 compsopogon
730	27	65.9	380	16	Q8E6X6	Q8e6x6 streptococc
731	27	65.9	380	16	Q8E1F9	Q8elf9 streptococc
732	27	65.9	382	2	Q9RM37	Q9rm37 escherichia
733	27	65.9	382	10	Q9LHW0	Q9lhw0 oryza sativ
734	27	65.9	383	5	Q8I280	Q8i280 plasmodium
735	27	65.9	383	10	Q9SKP8	Q9skp8 arabidopsis
736	27	65.9	384	8	Q8LWZ3	Q8lwz3 laminaria d
737	27	65.9	384	16	Q8ESE3	Q8ese3 oceanobacil
738	27	65.9	385	16	Q9AA68	Q9aa68 caulobacter
739	27	65.9	389	16	Q89UG7	Q89ug7 bradyrhizob
740	27	65.9	390	5	Q9TXJ3	Q9txj3 leishmania
741	27	65.9	390	16	Q8CMV2	Q8cmv2 staphylococ

742	27	65.9	393	10	Q9ZWA1	Q9zwa1 arabidopsis
743	27	65.9	393	10	Q8GZ44	Q8gz44 arabidopsis
744	27	65.9	393	16	Q9A5N6	Q9a5n6 caulobacter
745	27	65.9	394	16	P73566	P73566 synechocyst
746	27	65.9	398	5	Q26004	Q26004 plasmodium
747	27	65.9	398	5	Q9U433	Q9u433 plasmodium
748	27	65.9	398	5	Q9U432	Q9u432 plasmodium
749	27	65.9	398	5	Q9TVT9	Q9tvt9 plasmodium
750	27	65.9	398	5	Q8I484	Q8i484 plasmodium
751	27	65.9	398	12	Q8BA41	Q8ba41 avian param
752	27	65.9	399	16	Q9PN56	Q9pn56 campylobact
753	27	65.9	400	5	Q8MUN1	Q8mun1 plasmodium
754	27	65.9	400	5	Q8I485	Q8i485 plasmodium
755	27	65.9	400	16	Q8YKU6	Q8yku6 anabaena sp
756	27	65.9	403	5	Q8T2W3	Q8t2w3 trypanosoma
757	27	65.9	405	8	Q8MAA3	Q8maa3 rhodochaete
758	27	65.9	405	10	Q9SXC4	Q9sxc4 arabidopsis
759	27	65.9	406	16	Q82SE8	Q82se8 nitrosomona
760	27	65.9	407	16	Q9ZDJ2	Q9z dj2 rickettsia
761	27	65.9	409	16	Q7VNU4	Q7vnu4 haemophilus
762	27	65.9	410	16	Q9RGT4	Q9rgt4 salmonella
763	27	65.9	410	17	Q979W6	Q979w6 thermoplasma
764	27	65.9	412	10	Q9SHS5	Q9shs5 arabidopsis
765	27	65.9	414	17	Q8U2N0	Q8u2n0 pyrococcus
766	27	65.9	415	12	P87751	P87751 hepatitis c
767	27	65.9	415	16	Q92S40	Q92s40 rhizobium m
768	27	65.9	417	16	O25454	O25454 helicobacte
769	27	65.9	418	16	Q8EHV4	Q8ehv4 shewanella
770	27	65.9	420	16	Q9PEW9	Q9pew9 xylella fas
771	27	65.9	420	16	Q87AP7	Q87ap7 xylella fas
772	27	65.9	423	3	Q04429	Q04429 saccharomyc
773	27	65.9	424	3	O74508	O74508 schizosacch
774	27	65.9	428	5	Q8SXA2	Q8sxa2 drosophila
775	27	65.9	430	5	Q17814	Q17814 caenorhabdi
776	27	65.9	431	8	Q31955	Q31955 compsopogon
777	27	65.9	431	16	O66787	O66787 aquifex ae
778	27	65.9	438	5	Q9VNC2	Q9vnc2 drosophila
779	27	65.9	438	16	Q8A8U0	Q8a8u0 bacteroides
780	27	65.9	439	4	Q9BUG9	Q9bug9 homo sapien
781	27	65.9	439	10	Q9C865	Q9c865 arabidopsis
782	27	65.9	441	16	O67719	O67719 aquifex ae
783	27	65.9	441	16	Q89RS5	Q89rs5 bradyrhizob
784	27	65.9	442	11	Q8C3Z6	Q8c3z6 mus musculu
785	27	65.9	442	11	Q921V5	Q921v5 mus musculu
786	27	65.9	444	11	Q9CV28	Q9cv28 mus musculu
787	27	65.9	445	4	Q9H8M7	Q9h8m7 homo sapien
788	27	65.9	445	5	Q9U2K7	Q9u2k7 caenorhabdi
789	27	65.9	446	6	O19071	O19071 s alpha-1,6
790	27	65.9	446	16	Q8ZfV5	Q8zfv5 yersinia pe
791	27	65.9	446	16	Q83A43	Q83a43 coxiella bu
792	27	65.9	447	2	Q93HR7	Q93hr7 vibrio para
793	27	65.9	451	10	Q9FNU6	Q9fnu6 populus x c
794	27	65.9	452	16	Q81V86	Q81v86 bacillus an
795	27	65.9	454	10	Q9FUE7	Q9fue7 populus del
796	27	65.9	454	10	Q9FUE9	Q9fue9 populus tri
797	27	65.9	454	10	Q9FUE8	Q9fue8 populus tri
798	27	65.9	455	4	Q96S13	Q96s13 homo sapien

799	27	65.9	455	5	Q8SXR5	Q8sxr5 drosophila
800	27	65.9	455	16	Q8CZE5	Q8cze5 oceanobacil
801	27	65.9	458	16	Q8DDA4	Q8dda4 vibrio vuln
802	27	65.9	459	16	Q83AP6	Q83ap6 coxiella bu
803	27	65.9	461	10	Q98RU1	Q98rul guillardia
804	27	65.9	463	10	Q7X8X9	Q7x8x9 oryza sativ
805	27	65.9	468	3	Q9P8W6	Q9p8w6 candida alb
806	27	65.9	470	2	Q51847	Q51847 prochloroth
807	27	65.9	470	4	Q9BZE6	Q9bze6 homo sapien
808	27	65.9	470	16	Q8YFJ3	Q8yfy3 brucella me
809	27	65.9	470	16	Q8G2B8	Q8g2b8 brucella su
810	27	65.9	470	17	Q8TQM4	Q8tqm4 methanosarc
811	27	65.9	473	2	Q9ZB35	Q9zb35 hydrogenoph
812	27	65.9	473	5	Q9VN12	Q9vn12 drosophila
813	27	65.9	475	8	Q8WJ20	Q8wj20 ristantia g
814	27	65.9	476	12	Q8QNG3	Q8qng3 ectocarpus
815	27	65.9	478	16	Q8YM51	Q8ym51 anabaena sp
816	27	65.9	480	17	O29235	O29235 archaeoglob
817	27	65.9	483	12	Q82172	Q82172 leporid her
818	27	65.9	485	5	Q18544	Q18544 caenorhabdi
819	27	65.9	488	2	O30662	O30662 vibrio chol
820	27	65.9	488	16	Q9KQ66	Q9kq66 vibrio chol
821	27	65.9	489	16	Q8XPS9	Q8xps9 ralstonia s
822	27	65.9	493	10	Q7XKG0	Q7xkg0 oryza sativ
823	27	65.9	501	16	Q7UZT9	Q7uzt9 prochloroco
824	27	65.9	503	2	Q8GJK0	Q8gjk0 meiothermus
825	27	65.9	503	8	Q8W797	Q8w797 lysicarpus
826	27	65.9	503	8	Q9TKD2	Q9tkd2 tristaniops
827	27	65.9	503	8	Q7YN83	Q7yn83 diospyros v
828	27	65.9	505	13	Q8AWC7	Q8awc7 gallus gall
829	27	65.9	505	16	O69566	O69566 mycobacteri
830	27	65.9	506	3	Q9USN0	Q9usn0 schizosacch
831	27	65.9	506	16	Q837F7	Q837f7 enterococcu
832	27	65.9	511	16	Q8CXK8	Q8cxk8 oceanobacil
833	27	65.9	513	5	Q9V8T4	Q9v8t4 drosophila
834	27	65.9	513	5	Q8IHB1	Q8ihb1 drosophila
835	27	65.9	514	5	O76958	O76958 leishmania
836	27	65.9	521	16	Q92DS3	Q92ds3 listeria in
837	27	65.9	522	16	Q8PKK8	Q8pkk8 xanthomonas
838	27	65.9	523	16	Q9CLU9	Q9clu9 pasteurella
839	27	65.9	530	11	Q8C2Z3	Q8c2z3 mus musculu
840	27	65.9	532	2	Q8KQZ4	Q8kqz4 vibrio chol
841	27	65.9	537	4	Q9HA80	Q9ha80 homo sapien
842	27	65.9	541	10	Q851K4	Q851k4 oryza sativ
843	27	65.9	542	10	Q9LYP8	Q9lyp8 arabidopsis
844	27	65.9	555	2	Q8KY42	Q8ky42 streptomyce
845	27	65.9	555	16	Q8U8S0	Q8u8s0 agrobacteri
846	27	65.9	557	10	Q9XIC5	Q9xic5 arabidopsis
847	27	65.9	568	5	Q9NAL4	Q9nal4 caenorhabdi
848	27	65.9	568	16	Q92HM4	Q92hm4 rickettsia
849	27	65.9	569	2	Q93CY6	Q93cy6 lactobacill
850	27	65.9	571	16	Q8XHF8	Q8xhf8 clostridium
851	27	65.9	576	16	O06137	O06137 mycobacteri
852	27	65.9	576	16	Q7TZV7	Q7tzv7 mycobacteri
853	27	65.9	579	10	Q8S8Q2	Q8s8q2 arabidopsis
854	27	65.9	580	16	Q9X2G1	Q9x2g1 thermotoga
855	27	65.9	581	16	Q8A7H6	Q8a7h6 bacteroides

856	27	65.9	582	10	Q23492	Q23492 arabidopsis
857	27	65.9	584	16	Q55449	Q55449 synechocyst
858	27	65.9	590	10	Q9M3B7	Q9m3b7 arabidopsis
859	27	65.9	594	12	Q9PYU0	Q9pyu0 xestia c-ni
860	27	65.9	597	10	004659	004659 arabidopsis
861	27	65.9	600	16	Q8E3U4	Q8e3u4 streptococc
862	27	65.9	600	16	Q8DY79	Q8dy79 streptococc
863	27	65.9	602	10	Q8LHS7	Q8lhs7 oryza sativ
864	27	65.9	602	16	Q92HU2	Q92hu2 rickettsia
865	27	65.9	604	5	Q9U1E2	Q9ule2 leishmania
866	27	65.9	605	12	Q8QVE4	Q8qve4 eyach virus
867	27	65.9	607	5	Q9VVC1	Q9vvc1 drosophila
868	27	65.9	607	10	Q9FL94	Q9fl94 arabidopsis
869	27	65.9	607	10	Q8L622	Q8l622 arabidopsis
870	27	65.9	612	16	Q9I0E3	Q9i0e3 pseudomonas
871	27	65.9	612	16	Q8Y2U3	Q8y2u3 ralstonia s
872	27	65.9	613	16	Q8P8R4	Q8p8r4 xanthomonas
873	27	65.9	616	5	Q8I324	Q8i324 plasmodium
874	27	65.9	616	10	Q9LV10	Q9lv10 arabidopsis
875	27	65.9	617	2	Q8RP27	Q8rp27 actinomadur
876	27	65.9	625	2	Q93EB1	Q93eb1 rhizobium l
877	27	65.9	625	16	Q8U8V6	Q8u8v6 agrobacteri
878	27	65.9	627	16	Q89MD4	Q89md4 bradyrhizob
879	27	65.9	628	5	Q8IPM4	Q8ipm4 drosophila
880	27	65.9	628	16	Q89EW3	Q89ew3 bradyrhizob
881	27	65.9	631	16	Q9KE29	Q9ke29 bacillus ha
882	27	65.9	631	16	Q8EN36	Q8en36 oceanobacil
883	27	65.9	631	16	Q81YT2	Q81yt2 bacillus an
884	27	65.9	631	16	Q81I69	Q81i69 bacillus ce
885	27	65.9	639	16	Q8D8E9	Q8d8e9 vibrio vuln
886	27	65.9	640	16	Q92PE5	Q92pe5 rhizobium m
887	27	65.9	641	3	Q870Z9	Q870z9 neurospora
888	27	65.9	650	17	Q8TSE0	Q8tse0 methanosarc
889	27	65.9	659	2	Q8RP47	Q8rp47 azospirillu
890	27	65.9	660	16	Q930J2	Q930j2 rhizobium m
891	27	65.9	660	16	Q88CC9	Q88cc9 pseudomonas
892	27	65.9	661	16	Q92Y86	Q92y86 rhizobium m
893	27	65.9	662	5	Q9VLK1	Q9vlk1 drosophila
894	27	65.9	667	16	Q7UZL0	Q7uzl0 prochloroco
895	27	65.9	669	17	Q8PVI7	Q8pvi7 methanosarc
896	27	65.9	694	16	Q97T45	Q97t45 streptococc
897	27	65.9	702	16	Q8UD86	Q8ud86 agrobacteri
898	27	65.9	702	16	Q8DRI8	Q8dri8 streptococc
899	27	65.9	706	5	Q9VYR5	Q9vyr5 drosophila
900	27	65.9	736	5	Q44857	Q44857 caenorhabdi
901	27	65.9	743	13	Q9YGE7	Q9yge7 oncorhynch
902	27	65.9	749	12	Q7T4K3	Q7t4k3 pea stem ne
903	27	65.9	750	11	Q80T69	Q80t69 mus musculu
904	27	65.9	754	4	Q9NUP6	Q9nup6 homo sapien
905	27	65.9	761	2	Q9F7S9	Q9f7s9 gamma-prote
906	27	65.9	779	4	Q9H3P4	Q9h3p4 homo sapien
907	27	65.9	780	17	Q9HJC5	Q9hjc5 thermoplasm
908	27	65.9	782	16	Q8DBS6	Q8dbs6 vibrio vuln
909	27	65.9	783	16	Q9A842	Q9a842 caulobacter
910	27	65.9	784	10	Q9FVA3	Q9fva3 zea mays (m
911	27	65.9	791	16	Q98RJ1	Q98rj1 mycoplasma
912	27	65.9	795	11	Q7TNJ3	Q7tnj3 mus musculu

913	27	65.9	826	4	Q8IY15	Q8iy15 homo sapien
914	27	65.9	840	2	Q8KLI9	Q8kli9 thermomonos
915	27	65.9	842	16	Q8PN87	Q8pn87 xanthomonas
916	27	65.9	861	16	Q8PBN1	Q8pbn1 xanthomonas
917	27	65.9	865	5	P90740	P90740 caenorhabdi
918	27	65.9	879	2	Q8KKY1	Q8kkyl rhizobium e
919	27	65.9	879	2	O30579	O30579 bacillus su
920	27	65.9	879	16	O31504	O31504 bacillus su
921	27	65.9	884	5	Q8I7A1	Q8i7a1 plasmodium
922	27	65.9	895	5	Q86KM1	Q86km1 dictyosteli
923	27	65.9	895	5	Q7Z1B9	Q7z1b9 plasmodium
924	27	65.9	915	16	Q7VCW0	Q7vcw0 prochloroco
925	27	65.9	923	10	Q9FTZ7	Q9ftz7 oryza sativ
926	27	65.9	925	5	Q9U7E8	Q9u7e8 caenorhabdi
927	27	65.9	930	16	Q8E6N1	Q8e6n1 streptococc
928	27	65.9	930	16	Q8E177	Q8e177 streptococc
929	27	65.9	930	16	Q8DVD3	Q8dvd3 streptococc
930	27	65.9	930	16	Q8DNW4	Q8dnw4 streptococc
931	27	65.9	932	16	Q88V89	Q88v89 lactobacill
932	27	65.9	933	16	Q99YW3	Q99yw3 streptococc
933	27	65.9	933	16	Q8P068	Q8p068 streptococc
934	27	65.9	933	16	Q8K6S2	Q8k6s2 streptococc
935	27	65.9	949	3	O14093	O14093 schizosacch
936	27	65.9	951	13	Q7ZUF4	Q7zuf4 brachydanio
937	27	65.9	954	4	Q8WXV8	Q8wxv8 homo sapien
938	27	65.9	957	4	Q9H0V3	Q9h0v3 homo sapien
939	27	65.9	957	4	Q96P44	Q96p44 homo sapien
940	27	65.9	971	5	Q95Q20	Q95q20 caenorhabdi
941	27	65.9	996	16	Q97NR7	Q97nr7 streptococc
942	27	65.9	996	16	Q8DNF6	Q8dnf6 streptococc
943	27	65.9	1014	10	Q84KC5	Q84kc5 hordeum vul
944	27	65.9	1021	5	O15733	O15733 dictyosteli
945	27	65.9	1024	11	Q9ESJ5	Q9esj5 mus musculu
946	27	65.9	1025	16	Q830R3	Q830r3 enterococcu
947	27	65.9	1030	16	Q8DSP2	Q8dsp2 streptococc
948	27	65.9	1056	10	Q7XV85	Q7xv85 oryza sativ
949	27	65.9	1059	10	Q8LI36	Q8li36 oryza sativ
950	27	65.9	1061	13	Q9I9K0	Q9i9k0 xenopus lae
951	27	65.9	1074	16	Q88TG3	Q88tg3 lactobacill
952	27	65.9	1079	12	Q9YW56	Q9yw56 melanoplus
953	27	65.9	1105	5	Q9VX31	Q9vx31 drosophila
954	27	65.9	1106	5	Q8MMD3	Q8mmd3 drosophila
955	27	65.9	1117	5	Q9VK50	Q9vk50 drosophila
956	27	65.9	1121	5	Q9NJ94	Q9nj94 drosophila
957	27	65.9	1124	16	Q9CM76	Q9cm76 pasteurilla
958	27	65.9	1153	16	Q92JC1	Q92jc1 rickettsia
959	27	65.9	1155	16	Q9ZE40	Q9ze40 rickettsia
960	27	65.9	1165	5	Q962L9	Q962l9 plasmodium
961	27	65.9	1168	10	Q9LE81	Q9le81 arabidopsis
962	27	65.9	1182	11	Q8C6K9	Q8c6k9 mus musculu
963	27	65.9	1196	10	Q9SG02	Q9sg02 arabidopsis
964	27	65.9	1196	10	Q9LKP0	Q9lkp0 arabidopsis
965	27	65.9	1218	10	Q8LHH9	Q8lhh9 oryza sativ
966	27	65.9	1225	2	Q8RJI1	Q8rji1 providencia
967	27	65.9	1225	2	Q849Z9	Q849z9 escherichia
968	27	65.9	1232	5	Q8MMD2	Q8mmd2 drosophila
969	27	65.9	1253	5	Q8WQ61	Q8wq61 drosophila



970	27	65.9	1253	5	Q9W111	Q9w111 drosophila
971	27	65.9	1253	5	Q8IKA0	Q8ika0 plasmodium
972	27	65.9	1293	10	Q9M8Z0	Q9m8z0 arabidopsis
973	27	65.9	1309	10	Q9SXF0	Q9sxf0 arabidopsis
974	27	65.9	1321	10	O64418	O64418 arabidopsis
975	27	65.9	1431	5	Q9NG75	Q9ng75 parartemia
976	27	65.9	1457	10	Q8LPU6	Q8lpu6 zea mays (m
977	27	65.9	1466	10	Q94IU9	Q94iu9 arabidopsis
978	27	65.9	1470	11	Q8CHD4	Q8chd4 mus musculu
979	27	65.9	1470	12	Q8JYK1	Q8jyk1 fiji diseas
980	27	65.9	1488	4	Q9HCM3	Q9hcm3 homo sapien
981	27	65.9	1525	10	O65343	O65343 zea mays (m
982	27	65.9	1557	11	Q811F2	Q811f2 mus musculu
983	27	65.9	1620	5	O45000	O45000 caenorhabdi
984	27	65.9	1725	6	Q867A1	Q867a1 canis famil
985	27	65.9	1725	11	P70570	P70570 rattus norv
986	27	65.9	1743	4	Q9BYP7	Q9byp7 homo sapien
987	27	65.9	1800	4	Q8TCX6	Q8tcx6 homo sapien
988	27	65.9	1806	4	Q96TG0	Q96tg0 homo sapien
989	27	65.9	1820	12	Q90153	Q90153 ustilago ma
990	27	65.9	1840	11	Q60467	Q60467 cricetulus
991	27	65.9	1998	12	O56050	O56050 rhopalosiph
992	27	65.9	2181	5	Q8IFQ4	Q8ifq4 plasmodium
993	27	65.9	2201	5	Q8IFQ5	Q8ifq5 plasmodium
994	27	65.9	2219	4	Q9C0A3	Q9c0a3 homo sapien
995	27	65.9	2256	5	Q8I5L5	Q8i5l5 plasmodium
996	27	65.9	2267	5	Q8IBX1	Q8ibx1 plasmodium
997	27	65.9	2270	5	Q8I3E5	Q8i3e5 plasmodium
998	27	65.9	2296	4	Q9Y3S1	Q9y3s1 homo sapien
999	27	65.9	2613	5	Q9GYD1	Q9gyd1 leishmania
1000	27	65.9	3013	12	O92530	O92530 hepatitis c

# ALIGNMENTS

## RESULT 1

### Q9UCD1

ID Q9UCD1 PRELIMINARY; PRT; 28 AA.  
AC Q9UCD1;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Beta-amyloid peptide (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=94045685; PubMed=8229004;  
RA Vigo-Pelfrey C., Lee D., Keim P., Lieberburg I., Schenk D.B.;  
RT "Characterization of beta-amyloid peptide from human cerebrospinal  
RT fluid.";  
RL J. Neurochem. 61:1965-1968(1993).  
DR HSSP; P05067; IAMB.  
DR GO; GO:0016020; C:membrane; IEA.

DR InterPro; IPR001255; Beta-APP.  
DR Pfam; PF03494; Beta-APP; 1.  
SQ SEQUENCE 28 AA; 3244 MW; DE7BD081160AFC81 CRC64;

Query Match 85.4%; Score 35; DB 4; Length 28;  
Best Local Similarity 100.0%; Pred. No. 3.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
|||||||  
Db 17 LVFFAED 23

## RESULT 2

### Q9UCA9

ID Q9UCA9 PRELIMINARY; PRT; 30 AA.  
AC Q9UCA9;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Beta-amyloid protein (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=94153015; PubMed=8109908;  
RA Wisniewski T., Lalowski M., Levy E., Marques M.R., Frangione B.;  
RT "The amino acid sequence of neuritic plaque amyloid from a familial  
RT Alzheimer's disease patient."  
RL Ann. Neurol. 35:245-246(1994).  
DR HSSP; P05067; 1BA4.  
DR GO; GO:0016020; C:membrane; IEA.  
DR InterPro; IPR001255; Beta-APP.  
DR Pfam; PF03494; Beta-APP; 1.  
SQ SEQUENCE 30 AA; 3391 MW; FF4167ABD081160A CRC64;

Query Match 85.4%; Score 35; DB 4; Length 30;  
Best Local Similarity 100.0%; Pred. No. 3.7;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
|||||||  
Db 17 LVFFAED 23

## RESULT 3

### Q9UC33

ID Q9UC33 PRELIMINARY; PRT; 33 AA.  
AC Q9UC33;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Beta-amyloid peptide (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=93024877; PubMed=1406936;  
 RA Seubert P., Vigo-Pelfrey C., Esch F., Lee M., Dovey H., Davis D.,  
 RA Sinha S., Schlossmacher M., Whaley J., Swindlehurst C.;  
 RT "Isolation and quantification of soluble Alzheimer's beta-peptide from  
 RT biological fluids.";  
 RL Nature 359:325-327(1992).  
 DR HSSP; P05067; 1BA4.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR InterPro; IPR001255; Beta-APP.  
 DR Pfam; PF03494; Beta-APP; 1.  
 SQ SEQUENCE 33 AA; 3674 MW; B1DEFE2F4167ABD0 CRC64;

Query Match 85.4%; Score 35; DB 4; Length 33;  
 Best Local Similarity 100.0%; Pred. No. 4;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVFFAED 7  
 |||||  
 Db 17 LVFFAED 23

#### RESULT 4

O35463

ID O35463 PRELIMINARY; PRT; 79 AA.  
 AC O35463;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Alzheimer's amyloid beta protein (Fragment).  
 GN BETA APP.  
 OS Cricetulus griseus (Chinese hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Cricetulus.  
 OX NCBI\_TaxID=10029;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sambamurti K., Pinnix I., Gandhi S.;  
 RL Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF030413; AAB86608.1; -.  
 DR HSSP; P05067; 1BA4.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR InterPro; IPR001255; Beta-APP.  
 DR Pfam; PF03494; Beta-APP; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 79 79  
 SQ SEQUENCE 79 AA; 8538 MW; 37F2C6C3BFF3F597 CRC64;

Query Match 85.4%; Score 35; DB 11; Length 79;  
 Best Local Similarity 100.0%; Pred. No. 9.8;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVFFAED 7

|||||||  
Db 37 LVFFAED 43

RESULT 5

Q16020

ID Q16020 PRELIMINARY; PRT; 82 AA.  
AC Q16020;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Beta-amyloid peptide (Fragment).  
GN BETA APP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93236601; PubMed=8476439;  
RA Denman R.B., Rosenzwaig R., Miller D.L.;  
RT "A system for studying the effect(s) of familial Alzheimer disease  
RT mutations on the processing of the beta-amyloid peptide precursor."  
RL Biochem. Biophys. Res. Commun. 192:96-103(1993).  
DR EMBL; S61383; AAB26265.2; -.  
DR HSSP; P05067; 1BA4.  
DR GO; GO:0016020; C:membrane; IEA.  
DR InterPro; IPR001255; Beta-APP.  
DR Pfam; PF03494; Beta-APP; 1.  
FT NON\_TER 1 1  
FT NON\_TER 82 82  
SQ SEQUENCE 82 AA; 8882 MW; F534AA5AE5D9230A CRC64;

Query Match 85.4%; Score 35; DB 4; Length 82;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
|||||||  
Db 34 LVFFAED 40

RESULT 6

Q16014

ID Q16014 PRELIMINARY; PRT; 82 AA.  
AC Q16014;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Beta-amyloid peptide (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93236601; PubMed=8476439;

RA Denman R.B., Rosenzcwaig R., Miller D.L.;  
 RT "A system for studying the effect(s) of familial Alzheimer disease  
 mutations on the processing of the beta-amyloid peptide precursor.";  
 RL Biochem. Biophys. Res. Commun. 192:96-103(1993).  
 DR EMBL; S60721; AAB26263.2; -.  
 DR HSSP; P05067; 1BA4.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR InterPro; IPR001255; Beta-APP.  
 DR Pfam; PF03494; Beta-APP; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 82 82  
 SQ SEQUENCE 82 AA; 8972 MW; F534AA5B3EA9230A CRC64;  
  
 Query Match 85.4%; Score 35; DB 4; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 LVFFAED 7  
 |||||  
 Db 34 LVFFAED 40

# RESULT 7

Q16019  
 ID Q16019 PRELIMINARY; PRT; 82 AA.  
 AC Q16019;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Beta-amyloid peptide (Fragment).  
 GN BETA APP.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93236601; PubMed=8476439;  
 RA Denman R.B., Rosenzcwaig R., Miller D.L.;  
 RT "A system for studying the effect(s) of familial Alzheimer disease  
 mutations on the processing of the beta-amyloid peptide precursor.";  
 RL Biochem. Biophys. Res. Commun. 192:96-103(1993).  
 DR EMBL; S61380; AAB26264.2; -.  
 DR HSSP; P05067; 1BA4.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR InterPro; IPR001255; Beta-APP.  
 DR Pfam; PF03494; Beta-APP; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 82 82  
 SQ SEQUENCE 82 AA; 8938 MW; F534AA50E579230A CRC64;  
  
 Query Match 85.4%; Score 35; DB 4; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 LVFFAED 7  
 |||||

Db 34 LVFFAED 40

RESULT 8

Q8JH58

ID Q8JH58 PRELIMINARY; PRT; 113 AA.  
AC Q8JH58;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Amyloid beta protein (Fragment).  
OS Chelydra serpentina serpentina (common snapping turtle).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Testudines; Cryptodira; Testudinoidea; Chelydridae; Chelydra.  
OX NCBI\_TaxID=134619;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21876906; PubMed=11882478;  
RA Trudeau V.L., Chiu S., Kennedy S.W., Brooks R.J.;  
RT "Octylphenol (OP) alters the expression of members of the amyloid  
RT protein family in the hypothalamus of the snapping turtle, Chelydra  
RT serpentina serpentina.";  
RL Environ. Health Perspect. 110:269-275(2002).  
DR EMBL; AF541917; AAN04908.1; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR InterPro; IPR008155; A4\_APP.  
DR InterPro; IPR001255; Beta-APP.  
DR Pfam; PF03494; Beta-APP; 1.  
DR PRINTS; PR00203; AMYLOIDA4.  
DR PROSITE; PS00320; A4\_INTRA; 1.  
FT NON\_TER 1 1  
SQ SEQUENCE 113 AA; 12750 MW; 72515C930496E053 CRC64;

Query Match 85.4%; Score 35; DB 13; Length 113;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
| | | | | | | |  
Db 31 LVFFAED 37

RESULT 9

Q8BPV5

ID Q8BPV5 PRELIMINARY; PRT; 218 AA.  
AC Q8BPV5;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Amyloid beta (Fragment).  
GN APP.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Lung;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK052448; BAC34997.1; -.  
 DR MGD; MGI:88059; App.  
 DR GO; GO:0005515; F:protein binding; IPI.  
 DR InterPro; IPR008155; A4\_APP.  
 DR InterPro; IPR001255; Beta-APP.  
 DR Pfam; PF03494; Beta-APP; 1.  
 DR PRINTS; PR00203; AMYLOIDA4.  
 DR PROSITE; PS00320; A4\_INTRA; 1.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 218 AA; 24118 MW; 95B55AFDAE1D0EF5 CRC64;

Query Match 85.4%; Score 35; DB 11; Length 218;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
 |||||  
 Db 136 LVFFAED 142

# RESULT 10

Q8E547

ID Q8E547 PRELIMINARY; PRT; 295 AA.  
 AC Q8E547;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN GBS1185.  
 OS Streptococcus agalactiae (serotype III).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=216495;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NEM316 / Serotype III;  
 RX MEDLINE=22242508; PubMed=12354221;  
 RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,  
 RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,  
 RA Kunst F.;  
 RT "Genome sequence of Streptococcus agalactiae, a pathogen causing  
 RT invasive neonatal disease."  
 RL Mol. Microbiol. 45:1499-1513(2002).  
 DR EMBL; AL766849; CAD46844.1; -.  
 DR SagaList; gbs1185; -.  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005179; F:hormone activity; IEA.  
 DR InterPro; IPR000187; corticoliberin.  
 DR InterPro; IPR000620; DUF6.

DR InterPro; IPR004626; RarD.  
DR Pfam; PF00892; DUF6; 1.  
DR TIGRFAMs; TIGR00688; rarD; 1.  
DR PROSITE; PS00511; CRF; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 295 AA; 33015 MW; 60DDE324099DD314 CRC64;

Query Match 85.4%; Score 35; DB 16; Length 295;  
Best Local Similarity 75.0%; Pred. No. 38;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
:||||:|  
Db 196 IVFFAKDF 203

RESULT 11

Q8DZI3

ID Q8DZI3 PRELIMINARY; PRT; 295 AA.  
AC Q8DZI3;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE RarD protein.  
GN RARD OR SAG1118.  
OS Streptococcus agalactiae (serotype V).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=216466;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=2603 V/R / Serotype V;  
RX MEDLINE=22222988; PubMed=12200547;  
RA Tettelin H., Masignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,  
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,  
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,  
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,  
RA Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,  
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,  
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,  
RA Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,  
RA Fraser C.M.;  
RT "Complete genome sequence and comparative genomic analysis of an  
RT emerging human pathogen, serotype V Streptococcus agalactiae.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).  
DR EMBL; AE014243; AAM99999.1; -.  
DR TIGR; SAG1118; -.  
DR GO; GO:0005576; C:extracellular; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005179; F:hormone activity; IEA.  
DR InterPro; IPR000187; corticoliberin.  
DR InterPro; IPR000620; DUF6.  
DR InterPro; IPR004626; RarD.  
DR Pfam; PF00892; DUF6; 1.  
DR TIGRFAMs; TIGR00688; rarD; 1.  
DR PROSITE; PS00511; CRF; 1.  
KW Complete proteome.



SQ SEQUENCE 295 AA; 33015 MW; 60DDE324099DD314 CRC64;

Query Match 85.4%; Score 35; DB 16; Length 295;  
Best Local Similarity 75.0%; Pred. No. 38;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
:||||:|  
Db 196 IVFFAKDF 203

RESULT 12

Q8UUI8

ID Q8UUI8 PRELIMINARY; PRT; 357 AA.  
AC Q8UUI8;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Putative mebrane protein (Fragment).  
GN APPA.  
OS Brachydanio rerio (Zebrafish) (Danio rerio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RX PubMed=11862463;  
RA Musa A., Lehrach H., Russo V.E.A.;  
RT "Distinct expression patterns of two zebrafish homologues of the human  
RT APP gene during embryonic development."  
RL Dev. Genes Evol. 211:563-567(2001).  
DR EMBL; AJ315637; CAC85734.1; -.  
DR ZFIN; ZDB-GENE-000616-13; appa.  
DR GO; GO:0016020; C:membrane; IEA.  
DR InterPro; IPR008155; A4\_APP.  
DR InterPro; IPR001255; Beta-APP.  
DR Pfam; PF03494; Beta-APP; 1.  
DR PRINTS; PR00203; AMYLOIDA4.  
DR PROSITE; PS00320; A4\_INTRA; 1.  
FT NON\_TER 1 1  
SQ SEQUENCE 357 AA; 40962 MW; 07D99EEF6C55B2D8 CRC64;

Query Match 85.4%; Score 35; DB 13; Length 357;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
|||||||  
Db 275 LVFFAED 281

RESULT 13

O20025

ID O20025 PRELIMINARY; PRT; 361 AA.  
AC O20025;

DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Intron maturase (Maturase K) (Fragment).  
 GN MATK.  
 OS Crithmum maritimum (samphire).  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
 OC campanulids; Apiales; Apiaceae; Apioideae; apioid superclade;  
 OC Pyramidoptereae; Crithmum.  
 OX NCBI\_TaxID=40916;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Plunkett G.M., Soltis D.E., Soltis P.S.;  
 RT "Evolutionary patterns in Apiaceae: inferences based on matK sequence  
 RT data.";  
 RL Syst. Bot. 21:477-495(1996).  
 CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II  
 CC INTRONS (BY SIMILARITY).  
 CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,  
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY  
 CC MITOCHONDRIAL INTRONS.  
 DR EMBL; U58558; AAB66262.1; -.  
 DR GO; GO:0009507; C:chloroplast; IEA.  
 DR GO; GO:0006397; P:mRNA processing; IEA.  
 DR InterPro; IPR002866; MatK\_N.  
 DR Pfam; PF01824; MatK\_N; 1.  
 KW mRNA processing; Chloroplast.  
 FT NON\_TER 361 361  
 SQ SEQUENCE 361 AA; 42847 MW; 43A0657ED3134DEA CRC64;

Query Match 85.4%; Score 35; DB 8; Length 361;  
 Best Local Similarity 75.0%; Pred. No. 46;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 | : | | | |  
 Db 73 LIFFANDE 80

#### RESULT 14

O20011

ID O20011 PRELIMINARY; PRT; 361 AA.  
 AC O20011;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Intron maturase (Maturase K) (Fragment).  
 GN MATK.  
 OS Anthriscus sylvestris.  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
 OC campanulids; Apiales; Apiaceae; Apioideae; Scandiceae; Scandicinae;  
 OC Anthriscus.  
 OX NCBI\_TaxID=48027;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Plunkett G.M., Soltis D.E., Soltis P.S.;  
 RT "Evolutionary patterns in Apiaceae: inferences based on matK sequence  
 RT data.";  
 RL Syst. Bot. 21:477-495(1996).  
 CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II  
 CC INTRONS (BY SIMILARITY).  
 CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,  
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY  
 CC MITOCHONDRIAL INTRONS.  
 DR EMBL; U58547; AAB66255.1; -.  
 DR GO; GO:0009507; C:chloroplast; IEA.  
 DR GO; GO:0006397; P:mRNA processing; IEA.  
 DR InterPro; IPR002866; MatK\_N.  
 DR Pfam; PF01824; MatK N; 1.  
 KW mRNA processing; Chloroplast.  
 FT NON\_TER 361 361  
 SQ SEQUENCE 361 AA; 43334 MW; D1A875A9910B6F21 CRC64;

Query Match 85.4%; Score 35; DB 8; Length 361;  
 BestLocal Similarity 75.0%; Pred. No. 46;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 | : | | | | |  
 Db 73 LIFFANDF 80

# RESULT 15

O20068

ID O20068 PRELIMINARY; PRT; 364 AA.  
 AC O20068;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Intron maturase (Maturase K) (Fragment).  
 GN MATK.  
 OS Neogoezia minor.  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
 OC campanulids; Apiales; Apiaceae; Apioideae; Oenantheae; Neogoezia.  
 OX NCBI\_TaxID=46372;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Plunkett G.M., Soltis D.E., Soltis P.S.;  
 RT "Evolutionary patterns in Apiaceae: inferences based on matK sequence  
 RT data.";  
 RL Syst. Bot. 21:477-495(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Plunkett G.M., Soltis D.E., Soltis P.S.;  
 RT "Clarification of the relationship between Apiaceae and Araliaceae  
 RT based on matK and rbcL sequence data.";  
 RL Am. J. Bot. 84:565-580(1997).  
 CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II

CC INTRONS (BY SIMILARITY).  
 CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,  
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY  
 CC MITOCHONDRIAL INTRONS.  
 DR EMBL; U58570; AAB66281.1; -.  
 DR GO; GO:0009507; C:chloroplast; IEA.  
 DR GO; GO:0006397; P:mRNA processing; IEA.  
 DR InterPro; IPR002866; MatK\_N.  
 DR Pfam; PF01824; MatK\_N; 1.  
 KW mRNA processing; Chloroplast.  
 FT NON\_TER 364 364  
 SQ SEQUENCE 364 AA; 42939 MW; DBABC9499ED36646 CRC64;

Query Match 85.4%; Score 35; DB 8; Length 364;  
 Best Local Similarity 75.0%; Pred. No. 47;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 |:| | | |  
 Db 76 LIFFANDF 83

# RESULT 16

## Q8BPC7

ID Q8BPC7 PRELIMINARY; PRT; 384 AA.  
 AC Q8BPC7;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Amyloid beta (Fragment).  
 GN APP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK076506; BAC36369.1; -.  
 DR MGD; MGI:88059; App.  
 DR GO; GO:0005515; F:protein binding; IPI.  
 DR InterPro; IPR008155; A4\_APP.  
 DR InterPro; IPR001255; Beta-APP.  
 DR Pfam; PF03494; Beta-APP; 1.  
 DR PRINTS; PR00203; AMYLOIDA4.  
 DR PROSITE; PS00320; A4\_INTRA; 1.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 384 AA; 43990 MW; A81B1AD8AE683173 CRC64;

Query Match 85.4%; Score 35; DB 11; Length 384;  
 Best Local Similarity 100.0%; Pred. No. 49;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
 |||||

Db 302 LVFFAED 308

RESULT 17

Q8UUS0

ID Q8UUS0 PRELIMINARY; PRT; 472 AA.  
 AC Q8UUS0;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Putative membrane protein (Fragment).  
 GN APPA.  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX PubMed=11862463;  
 RA Musa A., Lehrach H., Russo V.E.A.;  
 RT "Distinct expression patterns of two zebrafish homologues of the human  
 RT APP gene during embryonic development."  
 RL Dev. Genes Evol. 211:563-567(2001).  
 DR EMBL; AJ315636; CAC85733.1; -.  
 DR ZFIN; ZDB-GENE-000616-13; appa.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR InterPro; IPR008155; A4\_APP.  
 DR InterPro; IPR001255; Beta-APP.  
 DR Pfam; PF03494; Beta-APP; 1.  
 DR PRINTS; PR00203; AMYLOIDA4.  
 DR PROSITE; PS00320; A4\_INTRA; 1.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 472 AA; 53787 MW; 24F7128BE3356550 CRC64;

Query Match 85.4%; Score 35; DB 13; Length 472;  
 Best Local Similarity 100.0%; Pred. No. 61;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
 |||||

Db 390 LVFFAED 396

RESULT 18

O93296

ID O93296 PRELIMINARY; PRT; 534 AA.  
 AC O93296;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Amyloid protein (Fragment).

OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98337885; PubMed=9671674;  
 RA Barnes N.Y., Li L., Yoshikawa K., Schwartz L.M., Oppenheim R.W.,  
 RA Milligan C.E.;  
 RT "Increased production of amyloid precursor protein provides a  
 RT substrate for caspase-3 in dying motoneurons."  
 RL J. Neurosci. 18:5869-5880(1998).  
 DR EMBL; AF042098; AAC25052.1; -.  
 DR HSSP; P05067; 1BA4.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR InterPro; IPR008155; A4\_APP.  
 DR InterPro; IPR008154; A4\_extra.  
 DR InterPro; IPR001255; Beta-APP.  
 DR Pfam; PF02177; A4\_EXTRA; 1.  
 DR Pfam; PF03494; Beta-APP; 1.  
 DR PRINTS; PR00203; AMYLOIDA4.  
 DR PROSITE; PS00319; A4\_EXTRA; 1.  
 DR PROSITE; PS00320; A4\_INTRA; 1.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 534 AA; 60597 MW; FB53ECC2E66D4C92 CRC64;

Query Match 85.4%; Score 35; DB 13; Length 534;  
 Best Local Similarity 100.0%; Pred. No. 69;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
 |||||  
 Db 452 LVFFAED 458

#### RESULT 19

##### Q9PVL1

ID Q9PVL1 PRELIMINARY; PRT; 569 AA.  
 AC Q9PVL1;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Amyloid protein (Fragment).  
 GN APP.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Coulson E.J., Paliga K., Beyreuther K., Masters C.L.;  
 RT "What the evolution of the amyloid protein precursor supergene family  
 RT tells us about its function."  
 RL Neurochem. Int. 0:0-0(2000).

DR EMBL; AF030341; AAF12698.1; -.  
 DR HSSP; P05067; 1BA4.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR InterPro; IPR008155; A4\_APP.  
 DR InterPro; IPR008154; A4\_extra.  
 DR InterPro; IPR001255; Beta-APP.  
 DR Pfam; PF02177; A4\_EXTRA; 1.  
 DR Pfam; PF03494; Beta-APP; 1.  
 DR PRINTS; PR00203; AMYLOIDA4.  
 DR PROSITE; PS00319; A4\_EXTRA; 1.  
 DR PROSITE; PS00320; A4\_INTRA; 1.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 569 AA; 64753 MW; 0AB8BB851863A19D CRC64;

Query Match 85.4%; Score 35; DB 13; Length 569;  
 Best Local Similarity 100.0%; Pred. No. 74;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
 |||||  
 Db 488 LVFFAED 494

# RESULT 20

Q9I9E7

ID Q9I9E7 PRELIMINARY; PRT; 612 AA.  
 AC Q9I9E7;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Amyloid protein (Fragment).  
 GN APPA.  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Slavov D.B., Gardiner K.;  
 RT "An App cDNA from Zebrafish (Danio rerio).";  
 RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF257742; AAF71748.1; -.  
 DR HSSP; P05067; 1HZ3.  
 DR ZFIN; ZDB-GENE-000616-13; appa.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR InterPro; IPR008155; A4\_APP.  
 DR InterPro; IPR008154; A4\_extra.  
 DR InterPro; IPR001255; Beta-APP.  
 DR Pfam; PF02177; A4\_EXTRA; 1.  
 DR Pfam; PF03494; Beta-APP; 1.  
 DR PRINTS; PR00203; AMYLOIDA4.  
 DR PROSITE; PS00319; A4\_EXTRA; 1.  
 DR PROSITE; PS00320; A4\_INTRA; 1.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 612 AA; 69710 MW; 59A9ACBDF9C59EFF CRC64;

Query Match 85.4%; Score 35; DB 13; Length 612;  
Best Local Similarity 100.0%; Pred. No. 79;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
|||||||  
Db 530 LVFFAED 536

RESULT 21

Q7ZZT1

ID Q7ZZT1 PRELIMINARY; PRT; 678 AA.  
AC Q7ZZT1;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Amyloid protein a variant 2.  
GN APPA.  
OS Brachydanio rerio (Zebrafish) (Danio rerio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Groth C., Lardelli M.;  
RT "Investigation of zebrafish appa expression during embryogenesis.";  
RL Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AY271746; AAP22958.1; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR InterPro; IPR008155; A4\_APP.  
DR InterPro; IPR008154; A4\_extra.  
DR InterPro; IPR001255; Beta-APP.  
DR Pfam; PF02177; A4\_EXTRA; 1.  
DR Pfam; PF03494; Beta-APP; 1.  
DR PRINTS; PR00203; AMYLOIDA4.  
DR SMART; SM00006; A4\_EXTRA; 1.  
DR PROSITE; PS00319; A4\_EXTRA; 1.  
DR PROSITE; PS00320; A4\_INTRA; 1.  
SQ SEQUENCE 678 AA; 76755 MW; 94163778444FD0BC CRC64;

Query Match 85.4%; Score 35; DB 13; Length 678;  
Best Local Similarity 100.0%; Pred. No. 88;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
|||||||  
Db 596 LVFFAED 602

RESULT 22

Q9DGJ8

ID Q9DGJ8 PRELIMINARY; PRT; 695 AA.  
AC Q9DGJ8;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)



DE Beta-amyloid precursor protein 695 isoform.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sarasa M., Rodolosse A., Sorribas V.;  
 RT "Cloning of full-length chicken beta-amyloid precursor protein  
 RT isoforms.";  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF289218; AAG00593.1; -.  
 DR HSSP; P05067; 1BA4.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR InterPro; IPR008155; A4\_APP.  
 DR InterPro; IPR008154; A4\_extra.  
 DR InterPro; IPR001255; Beta-APP.  
 DR Pfam; PF02177; A4\_EXTRA; 1.  
 DR Pfam; PF03494; Beta-APP; 1.  
 DR PRINTS; PR00203; AMYLOIDA4.  
 DR SMART; SM00006; A4\_EXTRA; 1.  
 DR PROSITE; PS00319; A4\_EXTRA; 1.  
 DR PROSITE; PS00320; A4\_INTRA; 1.  
 SQ SEQUENCE 695 AA; 78565 MW; F201ED02AEC86D95 CRC64;

Query Match 85.4%; Score 35; DB 13; Length 695;  
 Best Local Similarity 100.0%; Pred. No. 90;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVFFAED 7  
 |||||  
 Db 613 LVFFAED 619

# RESULT 23

Q90W28

ID Q90W28 PRELIMINARY; PRT; 738 AA.  
 AC Q90W28;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Amyloid precursor protein.  
 GN APPA OR APP.  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Groth C., Lardelli M.;  
 RT "Expression analysis of zebrafish app.";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF389401; AAK64495.1; -.  
 DR ZFIN; ZDB-GENE-000616-13; appa.  
 DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0004867; F:serine protease inhibitor activity; IEA.  
 DR InterPro; IPR008155; A4\_APP.  
 DR InterPro; IPR008154; A4\_extra.  
 DR InterPro; IPR001255; Beta-APP.  
 DR InterPro; IPR002223; Kunitz\_BPTI.  
 DR Pfam; PF02177; A4\_EXTRA; 1.  
 DR Pfam; PF03494; Beta-APP; 1.  
 DR Pfam; PF00014; Kunitz\_BPTI; 1.  
 DR PRINTS; PR00203; AMYLOIDA4.  
 DR PRINTS; PR00759; BASICPTASE.  
 DR ProDom; PD000222; Kunitz\_BPTI; 1.  
 DR SMART; SM00006; A4\_EXTRA; 1.  
 DR SMART; SM00131; KU; 1.  
 DR PROSITE; PS00319; A4\_EXTRA; 1.  
 DR PROSITE; PS00320; A4\_INTRA; 1.  
 DR PROSITE; PS00280; BPTI\_KUNITZ\_1; 1.  
 DR PROSITE; PS50279; BPTI\_KUNITZ\_2; 1.  
 KW Protease inhibitor; Serine protease inhibitor.  
 SQ SEQUENCE 738 AA; 83577 MW; AF480F6D308FD298 CRC64;

Query Match 85.4%; Score 35; DB 13; Length 738;  
 Best Local Similarity 100.0%; Pred. No. 96;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
 |||||  
 Db 656 LVFFAED 662

# RESULT 24

## Q9DGJ7

ID Q9DGJ7 PRELIMINARY; PRT; 751 AA.  
 AC Q9DGJ7;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Beta-amyloid precursor protein 751 isoform.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sarasa M., Rodolosse A., Sorribas V.;  
 RT "Cloning of full-length chicken beta-amyloid precursor protein  
 RT isoforms."  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF289219; AAG00594.1; -.  
 DR HSSP; P05067; 1BA4.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0004867; F:serine protease inhibitor activity; IEA.  
 DR InterPro; IPR008155; A4\_APP.  
 DR InterPro; IPR008154; A4\_extra.  
 DR InterPro; IPR001255; Beta-APP.  
 DR InterPro; IPR002223; Kunitz\_BPTI.  
 DR Pfam; PF02177; A4\_EXTRA; 1.

DR Pfam; PF03494; Beta-APP; 1.  
 DR Pfam; PF00014; Kunitz\_BPTI; 1.  
 DR PRINTS; PR00203; AMYLOIDA4.  
 DR PRINTS; PR00759; BASICPTASE.  
 DR ProDom; PD000222; Kunitz\_BPTI; 1.  
 DR SMART; SM00006; A4\_EXTRA; 1.  
 DR SMART; SM00131; KU; 1.  
 DR PROSITE; PS00319; A4\_EXTRA; 1.  
 DR PROSITE; PS00320; A4\_INTRA; 1.  
 DR PROSITE; PS00280; BPTI\_KUNITZ\_1; 1.  
 DR PROSITE; PS50279; BPTI\_KUNITZ\_2; 1.  
 KW Protease inhibitor; Serine protease inhibitor.  
 SQ SEQUENCE 751 AA; 84705 MW; E78E9413A8033D84 CRC64;

Query Match 85.4%; Score 35; DB 13; Length 751;  
 Best Local Similarity 100.0%; Pred. No. 98;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVFFAED 7  
 |||||  
 Db 669 LVFFAED 675

# RESULT 25

## Q8T9D3

ID Q8T9D3 PRELIMINARY; PRT; 1169 AA.  
 AC Q8T9D3;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE SD05642p.  
 GN CG7161.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanenvong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AY069816; AAL39961.1; -.  
 DR FlyBase; FBgn0035891; CG7161.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF00400; WD40; 4.  
 DR ProDom; PD000018; WD40; 1.  
 DR SMART; SM00320; WD40; 5.  
 DR PROSITE; PS50082; WD\_REPEATS\_2; 1.  
 DR PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
 KW Repeat; WD repeat.  
 SQ SEQUENCE 1169 AA; 131625 MW; 0FF6AC96B32B6C07 CRC64;

Query Match 85.4%; Score 35; DB 5; Length 1169;

Best Local Similarity 75.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
|:|:||||  
Db 1138 LLFYAEDF 1145

RESULT 26

Q9VSJ6

ID Q9VSJ6 PRELIMINARY; PRT; 1169 AA.  
AC Q9VSJ6;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE CG7161-PA.  
GN CG7161.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,

RA Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celniker S.E.,  
 RA Clamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J.,  
 RA Harris N.L., Kronmiller B., Marshall B., Millburn G.H., Richter J.,  
 RA Russo S., Searle S.M.J., Smith E., Shu S., Smutniak F.,  
 RA Whitfield E.J., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J.,  
 RA Lewis S.E.;  
 RT "Annotation of *Drosophila melanogaster* genome."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AE003555; AAF50423.3; -.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF00400; WD40; 3.  
 DR ProDom; PD000018; WD40; 1.  
 DR SMART; SM00320; WD40; 4.  
 DR PROSITE; PS50082; WD\_REPEATS\_2; 1.  
 DR PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
 SQ SEQUENCE 1169 AA; 131634 MW; 920176314C01C37D CRC64;

Query Match 85.4%; Score 35; DB 5; Length 1169;  
 Best Local Similarity 75.0%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 |::|||  
 Db 1138 LLFYAEDF 1145

# RESULT 27

Q21915

ID Q21915 PRELIMINARY; PRT; 222 AA.  
 AC Q21915;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE R10H10.1 protein.  
 GN R10H10.1.  
 OS *Caenorhabditis elegans*.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; *Caenorhabditis*.  
 OX NCBI\_TaxID=6239;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bardill S.C.;  
 RL Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology."  
 RL Science 282:2012-2018(1998).  
 DR EMBL; Z70686; CAA94609.1; -.  
 DR PIR; T24151; T24151.  
 DR WormPep; R10H10.1; CE06292.  
 DR InterPro; IPR001075; NifU\_C.  
 DR Pfam; PF01106; NifU-like; 1.  
 DR ProDom; PD002830; NifU\_C; 1.  
 SQ SEQUENCE 222 AA; 24757 MW; ED232B974B03E8D0 CRC64;

Query Match 80.5%; Score 33; DB 5; Length 222;  
 Best Local Similarity 85.7%; Pred. No. 74;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
 ||| |||  
 Db 71 VFFGEDF 77

# RESULT 28

Q9XSI7

ID Q9XSI7 PRELIMINARY; PRT; 261 AA.  
 AC Q9XSI7;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE MT1-MMP protein (Fragment).  
 GN MMP-14.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skeletal muscle;  
 RX MEDLINE=21071388; PubMed=11204721;  
 RA Balcerzak D., Querenguesser L., Dixon W.T., Baracos V.E.;  
 RT "Coordinate expression of matrix-degrading proteinases and their  
 RT activators and inhibitors in bovine skeletal muscle."  
 RL J. Anim. Sci. 79:94-107(2001).  
 DR EMBL; AF144758; AAD30298.1; -.  
 DR HSSP; P39900; 1JIZ.  
 DR MEROPS; M10.014; -.  
 DR GO; GO:0005578; C:extracellular matrix; IEA.  
 DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

DR InterPro; IPR000585; Hemopexin.  
 DR InterPro; IPR006026; Peptidase\_M.  
 DR InterPro; IPR001818; Pept\_M10A\_M12B.  
 DR InterPro; IPR006025; Pept\_M\_Zn\_BS.  
 DR Pfam; PF00045; hemopexin; 1.  
 DR Pfam; PF00413; Peptidase\_M10; 1.  
 DR PRINTS; PR00138; MATRIXIN.  
 DR SMART; SM00235; ZnMc; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Hydrolase; Metal-binding; Metalloprotease; Protease; Zinc.  
 FT NON\_TER 1 1  
 FT NON\_TER 261 261  
 SQ SEQUENCE 261 AA; 29854 MW; A0BF7CA8EDCF801F CRC64;

Query Match 80.5%; Score 33; DB 6; Length 261;  
 Best Local Similarity 75.0%; Pred. No. 87;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 | : | | | | |  
 Db 87 LIFFAEGF 94

# RESULT 29

Q87M09

ID Q87M09 PRELIMINARY; PRT; 448 AA.  
 AC Q87M09;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Na+-coupled multidrug efflux pump.  
 GN VP2449.  
 OS *Vibrio parahaemolyticus*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; *Vibrio*.  
 OX NCBI\_TaxID=670;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RIMD 2210633 / Serotype O3:K6;  
 RX MEDLINE=22508454; PubMed=12620739;  
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,  
 RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,  
 RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;  
 RT "Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism  
 RT distinct from that of *V. cholerae*."  
 RL Lancet 361:743-749(2003).  
 DR EMBL; AP005081; BAC60712.1; -.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0015297; F:antiporter activity; IEA.  
 DR GO; GO:0015238; F:drug transporter activity; IEA.  
 DR GO; GO:0006855; P:multidrug transport; IEA.  
 DR InterPro; IPR002528; MatE.  
 DR Pfam; PF01554; MatE; 2.  
 KW Complete proteome.  
 SQ SEQUENCE 448 AA; 48993 MW; 6B301E64FDB6BCC6 CRC64;

Query Match 80.5%; Score 33; DB 16; Length 448;

Best Local Similarity 75.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
| |||:|  
Db 105 LFFFADDF 112

RESULT 30

Q50563

ID Q50563 PRELIMINARY; PRT; 455 AA.  
AC Q50563;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Phosphomannomutase.  
GN MTH1584.  
OS Methanobacterium thermoautotrophicum.  
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
OC Methanobacteriaceae; Methanothermobacter.  
OX NCBI\_TaxID=187420;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Delta H;  
RX MEDLINE=98037514; PubMed=9371463;  
RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,  
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,  
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,  
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
RT deltaH: functional analysis and comparative genomics."  
RL J. Bacteriol. 179:7135-7155(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Delta H;  
RA Smith D.R.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.  
RN [3]  
RP SEQUENCE OF 1-174 FROM N.A.  
RC STRAIN=Delta H;  
RX MEDLINE=97158688; PubMed=9006048;  
RA Noelling J., Reeve J.N.;  
RT "Growth- and substrate-dependent transcription of the formate  
RT dehydrogenase (fdhCAB) operon in Methanobacterium thermoformicicum Z-  
RT 245."  
RL J. Bacteriol. 179:899-908(1997).  
DR EMBL; AE000918; AAB86057.1; -.  
DR EMBL; U51624; AAC44809.1; -.  
DR PIR; D69078; D69078.  
DR GO; GO:0016868; F:intramolecular transferase activity, phosph. . .; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR005841; PG/PMM\_mutase.  
DR InterPro; IPR005844; PG\_PMM\_ABAI.  
DR InterPro; IPR005845; PG\_PMM\_ABAIL.



DR InterPro; IPR005846; PG\_PMM\_ABABIII.  
 DR InterPro; IPR005843; PG\_PMM\_C.  
 DR Pfam; PF00408; PGM\_PMM; 1.  
 DR Pfam; PF02878; PGM\_PMM\_I; 1.  
 DR Pfam; PF02879; PGM\_PMM\_II; 1.  
 DR Pfam; PF02880; PGM\_PMM\_III; 1.  
 DR PRINTS; PR00509; PGMPMM.  
 DR PROSITE; PS00710; PGM\_PMM; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 455 AA; 49579 MW; E98E44E9F8A5C40B CRC64;

Query Match 80.5%; Score 33; DB 17; Length 455;  
 Best Local Similarity 71.4%; Pred. No. 1.5e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
 :||:|  
 Db 124 IFFSEDF 130

# RESULT 31

O62511

ID O62511 PRELIMINARY; PRT; 502 AA.  
 AC O62511;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE ZK550.2 protein.  
 GN ZK550.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Basham V.M.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology."  
 RL Science 282:2012-2018(1998).  
 DR EMBL; Z82287; CAB05312.1; -.  
 DR PIR; T27908; T27908.  
 DR WormPep; ZK550.2; CE16733.  
 DR InterPro; IPR007114; MFS.  
 DR PROSITE; PS50850; MFS; 1.  
 SQ SEQUENCE 502 AA; 55666 MW; B1B98A5B34254056 CRC64;

Query Match 80.5%; Score 33; DB 5; Length 502;  
 Best Local Similarity 75.0%; Pred. No. 1.7e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 |||||:

Db 227 LVFFQEDY 234

RESULT 32

Q8THF4

ID Q8THF4 PRELIMINARY; PRT; 651 AA.  
AC Q8THF4;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hydantoinase.  
GN HYUA OR MA4563.  
OS Methanosarcina acetivorans.  
OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;  
OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.  
OX NCBI\_TaxID=2214;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C2A / ATCC 35395 / DSM 2834;  
RX MEDLINE=21929760; PubMed=11932238;  
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,  
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,  
RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,  
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,  
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,  
RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,  
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,  
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,  
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,  
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,  
RA Metcalf W.W., Birren B.;  
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic  
RT and physiological diversity."  
RL Genome Res. 12:532-542(2002).  
DR EMBL; AE011177; AAM07902.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR002821; Hydantoinase\_A.  
DR InterPro; IPR008040; Hydant\_A\_N.  
DR Pfam; PF01968; Hydantoinase\_A; 1.  
DR Pfam; PF05378; Hydant\_A\_N; 1.  
KW Complete proteome.  
SQ SEQUENCE 651 AA; 71402 MW; 11214F3C3737F79B CRC64;

Query Match 80.5%; Score 33; DB 17; Length 651;  
Best Local Similarity 75.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
| |||:|  
Db 486 LTFFAKDF 493

RESULT 33

Q968T7

ID Q968T7 PRELIMINARY; PRT; 3610 AA.  
AC Q968T7;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Chloroquine resistance marker protein.  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5833;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K1;  
 RA Li G.D., Bray P.G., Ward S.A.;  
 RT "Possible chloroquine-resistance gene in Plasmodium falciparum.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF282241; AAK49448.1; -.  
 SQ SEQUENCE 3610 AA; 432028 MW; F6CAE8DD71F906 CRC64;

Query Match 80.5%; Score 33; DB 5; Length 3610;  
 Best Local Similarity 71.4%; Pred. No. 1.3e+03;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
 :||:|  
 Db 554 IFFSEDF 560

RESULT 34  
 Q968T6

ID Q968T6 PRELIMINARY; PRT; 3620 AA.  
 AC Q968T6;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Chloroquine resistance marker protein.  
 OS Plasmodium falciparum (isolate Dd2).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=57267;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Dd2;  
 RA Li G.D., Bray P.G., Ward S.A.;  
 RT "Possible chloroquine-resistance gene in Plasmodium falciparum.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF282242; AAK49449.1; -.  
 SQ SEQUENCE 3620 AA; 433151 MW; C6519F1F60F72349 CRC64;

Query Match 80.5%; Score 33; DB 5; Length 3620;  
 Best Local Similarity 71.4%; Pred. No. 1.3e+03;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
 :||:|  
 Db 553 IFFSEDF 559

RESULT 35  
 Q968Y0

ID Q968Y0 PRELIMINARY; PRT; 3628 AA.  
 AC Q968Y0;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Chloroquine resistance marker protein.  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5833;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Li G.D., Bray P.G., Ward S.A.;  
 RT "Plasmodium falciparum chloroquine resistance marker gene."  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF169404; AAK48418.1; -.  
 SQ SEQUENCE 3628 AA; 433967 MW; BCDBF5CDA3BA95FB CRC64;

Query Match 80.5%; Score 33; DB 5; Length 3628;  
 Best Local Similarity 71.4%; Pred. No. 1.3e+03;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
 :||:||||  
 Db 554 IFFSEDF 560

# RESULT 36

## Q8IKY8

ID Q8IKY8 PRELIMINARY; PRT; 3704 AA.  
 AC Q8IKY8;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Chloroquine resistance marker protein.  
 GN PF14\_0463.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=22255705; PubMed=12368864;  
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,  
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
 RA Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,  
 RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
 RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
 RA Fraser C.M., Barrell B.;  
 RT "Genome sequence of the human malaria parasite Plasmodium  
 RT falciparum."  
 RL Nature 419:498-511(2002).  
 DR EMBL; AE014823; AAN37076.1; -.  
 SQ SEQUENCE 3704 AA; 443067 MW; 27EF423D961E0B77 CRC64;

Query Match 80.5%; Score 33; DB 5; Length 3704;  
 Best Local Similarity 71.4%; Pred. No. 1.3e+03;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VFFAEDF 8  
:|:|  
Db 554 IFFSEDF 560

RESULT 37

Q9BHZ8

ID Q9BHZ8 PRELIMINARY; PRT; 60 AA.  
AC Q9BHZ8;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Thioredoxin peroxidase, putative (Fragment).  
GN THP1.  
OS Globodera rostochiensis (Golden nematode).  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
OC Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.  
OX NCBI\_TaxID=31243;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Qin L., Schots A., Smant G., Bakker J., Helder J.;  
RL Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AJ133420; CAC34452.1; -.  
DR GO; GO:0004601; F:peroxidase activity; IEA.  
DR InterPro; IPR000866; AhpC-TSA.  
DR Pfam; PF00578; AhpC-TSA; 1.  
KW Peroxidase.  
FT NON\_TER 1 1  
FT NON\_TER 60 60  
SQ SEQUENCE 60 AA; 6785 MW; 2BD9A80FFC7895B9 CRC64;

Query Match 78.0%; Score 32; DB 5; Length 60;  
Best Local Similarity 62.5%; Pred. No. 31;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
|:|:|  
Db 36 LIFYPEDF 43

RESULT 38

Q90160

ID Q90160 PRELIMINARY; PRT; 75 AA.  
AC Q90160;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
DE Chitinase.  
OS Bombyx mori nuclear polyhedrosis virus (BmNPV).  
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
OC Nucleopolyhedrovirus.  
OX NCBI\_TaxID=10458;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=T3;

RX MEDLINE=94365965; PubMed=8083997;  
 RA Ohkawa T., Majima K., Maeda S.;  
 RT "A cysteine protease encoded by the baculovirus Bombyx mori nuclear  
 RT polyhedrosis virus.";  
 RL J. Virol. 68:6619-6625(1994).  
 DR EMBL; U12688; AAB49543.1; -.  
 SQ SEQUENCE 75 AA; 8403 MW; A082D0827846701E CRC64;

Query Match 78.0%; Score 32; DB 12; Length 75;  
 Best Local Similarity 62.5%; Pred. No. 40;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVFFAEDF 8  
 |:|:| ||  
 Db 51 LIFYASDF 58

# RESULT 39

Q931V3

ID Q931V3 PRELIMINARY; PRT; 184 AA.  
 AC Q931V3;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE GTP pyrophosphokinase.  
 GN YJBM OR SAV1001.  
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699).  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxID=158878;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21311952; PubMed=11418146;  
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
 RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
 RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,  
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
 RT "Whole genome sequencing of meticillin-resistant Staphylococcus  
 RT aureus.";  
 RL Lancet 357:1225-1240(2001).  
 DR EMBL; AP003361; BAB57163.1; -.  
 DR GO; GO:0016301; F:kinase activity; IEA.  
 KW Kinase; Complete proteome.  
 SQ SEQUENCE 184 AA; 21755 MW; 347AD094AFCCCD03 CRC64;

Query Match 78.0%; Score 32; DB 16; Length 184;  
 Best Local Similarity 85.7%; Pred. No. 99;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVFFAED 7  
 ||||:||  
 Db 81 LVFFSED 87

# RESULT 40